DOCUMENT RESUME

ED 034 292 EA 002 602

AUTHOR	Strevell, Wallace H., Ed.
TITLE	Rationale of Education Evaluation.
INSTITUTION	Gulf Schools Supplementary Education Center, Pearland, Tex.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Elementary and Secondary Education.
REPORT NO	DPSC-66-1394; ESEA-Title-3
PUB DATE	15 Apr 67
NOTE	73p.; Position papers prepared by the Interdisciplinary Committee on Education Evaluation
EDRS PRICE	EDRS Price MF-$0.50 HC-$3.75
DESCRIPTORS	Administrator Role, Behavioral Science Research, Bibliographies, Board of Education Role, Community Control, Cultural Differences, *Evaluation Techniques, *Mathematical Models, Program Budgeting, *Program Evaluation, *Research Methodology, Self Actualization, *Student Evaluation, Student Subcultures, Systems Analysis, Teacher Role
IDENTIFIERS	ESEA Title III

ABSTRACT

This document, produced through an ESEA Title III grant, contains seven position papers on education evaluation from the standpoints of psychology, sociology, political science, business management, school administration, and computer science. The first paper summarizes the history of educational program evaluation and reviews current theories and techniques. The second discusses the implications of viewing evaluation as equivalent to, or a special case of, behavioral-science research. The third is concerned with subcultures of the school, self-actualization, and student evaluation. Evaluation of school-control is discussed in the fourth paper, with emphasis on dominant-submissive relationships which are customary in bureaucracies. The administrative function in evaluation is discussed in the fifth. The last two papers discuss program evaluation and the quantification of education for planning applications that use models for evaluation. An 87-entry bibliography is appended. (DE)

EA

ED034292

RATIONALE OF EDUCATION EVALUATION

Prepared by the

Interdisciplinary Committee on Education Evaluation

Wallace H. Strevell, EdD
Editor

GULF SCHOOLS SUPPLEMENTARY EDUCATION CENTER

Thomas Tope, Jr., Director
Assistant Directors
James A. Hefter
James C. McBride, EdD
Ralph O. Teter

Established through a grant under Title III of the Elementary and Secondary Education Act of 1965, P.L. 89-10

Pearland, Texas
1967

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EA 002 602

CONTENTS

Page

PREFACE. iii

CHAPTER

I. EVALUATION OF EDUCATIONAL PROGRAMS: AN EXPLORATION by Gene Atkinson and David A. Knickel 1

II. IMPLICATIONS OF VIEWING EDUCATIONAL EVALUATION AS RESEARCH IN BEHAVIORAL SCIENCES by Edwin P. Willems 10

III. SUBCULTURES OF THE SCHOOL; SELF-ACTUALIZATION AND STUDENT EVALUATION by Charles L. Mulford. 25

IV. EVALUATION OF SCHOOL CONTROL RELATIONSHIPS by Bancroft C. Henderson 31

V. EDUCATIONAL EVALUATION: ADMINISTRATIVE FUNCTION by Troy A. Womack 40

VI. PROGRAM EVALUATION IN EDUCATION by Wallace H. Strevell. . . . 48

VII. QUANTIFICATION OF EDUCATION FOR PLANNING APPLICATIONS UTILIZING MODELS FOR EVALUATION by Glen Self. 54

BIBLIOGRAPHY 64

PREFACE

The Interdisciplinary Committee on Education Evaluation was organized to explore the parameters of the evaluative function included in the charter of the Gulf Schools Supplementary Education Center (P.L. 89-10, Title III). The first stage of the committee's work has been to prepare position papers on education evaluation from the standpoint of psychology, sociology, political science, business management, school administration, and computer science.

Preliminary to the committee discussion, a study of the development of theory and practice in education evaluation and a bibliography were prepared by Dr. Gene Atkinson, Associate Professor in the Bureau of Education Research and Services, University of Houston, assisted by David A. Knickel. The position papers included in this report were produced by individuals on the committee and agreed to in a series of committee meetings.

The composition of the committee was designed to represent the viewpoints of different academic disciplines and institutions. It was thought that a task as complex and far-reaching as education evaluation should be of concern in many fields of endeavor. Accordingly, the committee was comprised of the following specialists:

Dr. Edwin P. Willems, Department of Psychology, Rice University

Dr. Charles Mulford, Department of Sociology, University of Houston

Dr. Bancroft C. Henderson, Department of Political Science, University of Houston

Mr. Troy A. Womack, Vice-President for Business Affairs, Houston Baptist College

Dr. Glen Self, Department of Industrial Engineering, Texas A & M University

Dr. Wallace H. Strevell, Department of Administrative Education, University of Houston - Chairman

The collection of papers should be read with appreciation of the contribution that each discipline has offered, since each author attempted to identify the research of his field that favored on education evaluation. The exercise, however, has had an essentially pragmatic purpose. We believe that guidelines have been stated for field testing a practical plan of evaluation.

W. H. Strevell, Chairman
Interdisciplinary Committee on
Education Evaluation

April 15, 1967

1

EVALUATION OF EDUCATIONAL PROGRAMS: AN EXPLORATION

Gene Atkinson, EdD*

> . . .evaluation has become a sort of sacred cow in education, in the sense that everyone, both educators and laymen, theoretically believes in it. Not many people know what it is or what to do with it, but it is nevertheless widely regarded as "a good thing." The vague notion is that we can find out through some procedure (scientific or otherwise) what we are accomplishing with all these federal dollars spent on education.
>
> --Harold Howe II, U. S.
> Commissioner of Education, December, 1966[1]

The above remark, introducing the U. S. Commissioner's views on evaluation while summarizing recent activities of his office, might well suggest caution to all who wish to evaluate educational programs. Whether or not "vague notions" and "procedures (scientific or otherwise)" in evaluation have in the past retarded educational progress, future practice requires the investment of the best available talent and techniques to insure maximum achievement of desirable objectives.

This report will briefly summarize the history of evaluation of educational programs and review current theories and techniques. Designed to give an introductory background to individuals interested in establishing or executing programs for educational evaluation, the report will include definitions and purposes of evaluation and will attempt to pinpoint major authors for those wishing to pursue specific topics in detail.

Definitions

Dressel states: "Evaluation involves judging the worth of an experience, idea, or process. The judgment presupposes standards or

---

*Associate Professor of Administrative Education, University of Houston, assisted by David A. Knickel, Graduate Research Assistant.

[1] Harold Howe II, "Growth and Growing Pains," Saturday Review, 49:68-70, 87, December 17, 1966.

criteria."[2] Shane's 1950 study showed five common interpretations by educators of the term "evaluation": (1) as the application of value to a problem (e.g., "criteria for evaluating books"), (2) as a synonym for measurement (testing), (3) as a label for processes used in gauging teacher resources, and (5) as a study of changes in children's behavior.[3] Only the latter two interpretations are considered here; both are incorporated in Tyler's "Basic Notions Regarding Evaluation":

> The process of evaluation is essentially the process of determining to what extent the educational objectives are actually being realized by the program of curriculum and instruction. However, since educational objectives are essentially changes in human beings, that is, the objectives aimed at are to produce certain desirable changes in the behavior patterns of the student, then evaluation is the process for determining the degree to which these changes in behavior are actually taking place.[4]

Travers, however, points out that evaluation is more than merely determining the actual outcomes (i.e., changes produced in the students); evaluation also involves a judgment of the desirability of whatever outcomes are demonstrated to occur.[5]

History

Evaluation of education has a history as long as that of education itself. Current systems of inquiry and analysis in the field share common roots, of course, with other fields of modern knowledge; the roots have been well traced in works on the history of modern science.

Experimental science has made phenomenal advances from the time of Galileo and Newton, but techniques for measuring and appraising human behavior became common only in the latter part of the nineteenth century.

---

[2] Paul L. Dressel and Associates, Evaluation in Higher Education (Cambridge: Riverside Press, 1961), p. 6.

[3] Harold G. Shane and E. T. McSwain, Evaluation and the Elementary Curriculum (New York: Henry Holt and Company, 1951), pp. 56-8.

[4] Ralph W. Tyler, Basic Principles of Curriculum and Instruction (Chicago: University of Chicago Press, 1950), p. 69.

[5] Robert M. W. Travers, Educational Measurement (New York: MacMillan Company, 1955), p. 7.

Three branches of the stream of science combined "to yield the vigorous measurement movement in psychology and its spread through education. . . These were (1) the physiological and experimental psychology that had its main growth in Germany in the nineteenth century, (2) Darwinian biology, and (3) the clinical concern for the maladjusted and underdeveloped individual."[6] Newly available quantitative data about humans and techniques for analyzing the data influenced research orientations not only in the subareas of psychology and education but in the other behavioral sciences as well. Although educators had appraised students and programs as best they could (generally by observation of student responses), new approaches were forthcoming; the measurement movement and new philosophical insights brought "wedges of inquiry and experimentation" into the thinking of the educational profession at the beginning of the twentieth century.[7]

One must note, in passing, the serious indictments by Callahan in Education and the Cult of Efficiency. Callahan argues persuasively, and with considerable evidence, that many of the school "evaluations" since 1900 have been less products of "inquiry and experimentation" than capitulation by educational leaders to pressures by business-industrial groups demanding lower costs and greater efficiencies," considering the schools basically as factories. Among the individuals receiving criticism were (alphabetically) Leonard Ayres, Franklin Bobbitt, Ellwood P. Cubberly, Fred Englehardt, Paul Hanus, Henry H. Linn, Frank Spaulding, and George Strayer. However, Callahan's accusation appears to have been directed less toward evaluative processes in general and more toward poor choices of objectives and criteria.[8]

Thomas has recently presented an input-output model of efficiency in education which attempts to meet Callahan's criticism by emphasizing that an efficient operation is "one which produces a maximum amount or level of the desired product, for given input of resources," with product defined broadly in terms of desirable educational goals.[9]

Since 1900, several major and hundreds of minor evaluative projects have been completed, many of these being oriented toward values other than economic efficiency (e.g., "quality" of education, individualized instruction, etc.). Accreditation of secondary schools by regional groups of colleges began in about 1900 (the College Entrance Examination Board was established in 1901).[10] The "Eight-Year Study" of the Progressive Education Association's Commission on the Relation of School and College utilized an elaborate and sophisticated evaluation program, with Ralph Tyler as Research Director.[11]

---

[6] Robert L. Thorndike and Elizabeth Hagen, Measurement and Evaluation in Psychology and Education (New York: John Wiley & Sons, Inc., 1961), p. 3.

[7] William Clark Trow, Scientific Method in Education (Boston: Houghton Mifflin Company, 1925), Introduction.

[8] Raymond E. Callahan, Education and the Cult of Efficiency (Chicago: University of Chicago Press, 1962).

[9] J. Alan Thomas, "Efficiency in Education" (Chicago: American Education Research Association Meeting, February 15, 1963). (Mimeographed.)

The Pennsylvania Study, conducted from about 1928-38 by Ben D. Wood and William S. Learned, set out "to review the current system of schooling in the light of objective evidence concerning its obtained results." Unfortunately, the especially developed "comprehensive" achievement tests apparently measured superficial knowledge of widely scattered facts rather than deep knowledge and thought processes. (Later Wood, working with the American Council on Education established the Cooperative Test Service while Learned, through the Carnegie Foundation, established the Graduate Record Examination.)[12]

The National Education Association's Committee on Tax Education and School finance cites several dozen studies which generally tend to support the Committee's conclusion that "better education does cost more." The report itself is a useful bibliographical guide.[13]

In 1956, the New York State Education Department created the Quality Measurement Project which began with a four-year exploratory study of one hundred school systems and 70,000 students, and has published a School Quality Workbook describing the current evaluative process in that state.[14]

Pennsylvania's General Assembly, in the School District Reorganization Act of 1963, required the State Board of Education to develop and to implement a plan for evaluating the quality of educational programs in the Commonwealth. Details of a proposed plan were submitted in a threevolume report in 1965.[15]

---

[10] Evaluative Criteria, 1960 Edition (Washington: National Study of Secondary School Evaluation, 1960), p. 3.

[11] Eugene R. Smith, Ralph W. Tyler and the Evaluation Staff, Appraising and Recording Student Progress (New York: Harper & Brothers, 1942).

[12] Travers, op. cit., pp. 79-82.

[13] National Education Association Committee on Tax Education and School Finance, Does Better Education Cost More? (Washington: The Association, 1959).

[14] William D. Firman, "The Quality Measurement Project in New York State," Science Education, 50:259-79, April, 1966.

Widespread interest in and concern with educational evaluation has been prompted recently by the Elementary and Secondary Education Act of 1965; statutory provisions require evaluation of all projects funded under provisions of this law.[16] The Congressional expectation of prompt and periodic evaluation of results elicited Howe's remarks cited earlier.[17]

Two nationwide evaluation programs which appear to be imminent are bringing about both improved thinking and heightened emotional reaction. The project, "Assessing the Progress of Education" (also known as the National Assessment Program), with Ralph Tyler as Chairman of the Exploratory Committee, aims to gather comprehensive, dependable, and valid data from which an index of educational progress, similar to the Gross National Product Index, can be derived. The educational progress index would afford a guide for the allocation of additional resources.[18] Pilot studies were scheduled for 1966, with inauguration of the full project to come in the fall of 1967. A separate proposal is the "projected cost-effectiveness study of the U. S. Office of Education planned to involve some 300,00 preschool children to college students in the fall of 1968. . ."[19]

Institutions of higher education have continually produced research-oriented evaluative studies, generally of themselves. The Cooperative Study of Evaluation in General Education of the American Council on Education sponsored development of techniques and measuring devices for college use.[20]

---

[15] Henry S. Dyer, "The Pennsylvania Plan, " Science Education, 50: 242-8, April, 1966.

[16] "How to Evaluate Federal Programs," Nation's Schools, 77:49-71, May, 1966.

[17] Harold Howe II, "Growth and Growing Pains," Saturday Review, 49:68-70, 87, December 17, 1966.

[18] Ralph W. Tyler, "Assessing the Progress of Education," Science Education, 50:239-42, April, 1966.

[19] J. Raymond Gerberich, "Assessment: A Forward Look," Educational Leadership, 24:115-8, November, 1966.

[20] Paul L. Dressel and Lewis B. Mayhew, General Education: Explorations in Evaluation (Washington: American Council on Education, 1954).

Hagen and Thorndike noted in 1957:

> . . .since World War II the focus of interest and effort appears to have shifted away from a research oriented attempt to develop new and better evaluation procedures, to an action research oriented attempt to involve school personnel in evaluating their own programs. . .This type of self-evaluation is defended not in terms of the technical excellence of the appraisal that will result, but on the ground that the evaluation experience will be a valuable learning experience for the teachers that engage in it.[21]

Recent literature, however, appears to indicate a strong resurgence of the research orientation.

Theory and Techniques

Evaluation involves "three distinct aspects: (a) selecting the attributes that are important for judging the worth of a specimen to be evaluated, (b) developing and applying procedures that will describe these attributes truly and accurately, and (c) synthesizing the evidence yielded by these procedures into a final judgment of worth."[22] Restated, these aspects become (a) identifying objectives, (b) developing and applying measuring instruments, and (c) interpreting results.

Objectives to be measured may be drawn from three domains, designated as structure, process, and product.[23] "Structure" includes such items as administrative data, school plant, and formal organization. Studies of financial support of schools, for instance, are of this type and tend to assume high correlation with desired outcomes. "Process" refers to some aspect of school or classroom procedure. Ryan reports a small research study of this type at the college level; Mort's Growing Edge is considered a process-oriented action research procedure for evaluating publis schools. Critics maintain that this type of study, too, generally can only assume high correlation with desired outcomes.[24,25,26] "Product" refers to the ultimate exhibited performance or behavior of students. The Eight-Year Study, the Quality Measurement Project in New York State, the current Pennsylvania Plan, and the proposed "National Assessment Program" represent product-oriented attempts at evaluation.

---

[21] Elizabeth P. Hagen and Robert L. Thorndike, "Evaluation," *Encyclopedia of Educational Research* (3rd ed.), pp. 482-6.

[22] *Ibid.*

[23] *Ibid.*

Selection of objectives to be measured depends on the judgment of the evaluator. Sources of objectives for evaluation are similar to and as varied as the sources of objectives for curriculum, discussed in standard works in the field of curriculum development.

Dressel points out that educational objectives may be examined relative to several continua: simple--complex, achievable--unachievable, explicit--implicit, intrinsic--transcendental, and goals for individual--goals for society.[27]

The Pennsylvania Plan has selected "ten major goals for education in Pennsylvania" as qualities to be measured; attempts are being made to translate these goals into measurable terms of student performance. The goals say that quality education

> . . .should help every child acquire (1) self-understanding and self-acceptance, (2) understanding and appreciation of social, cultural, and ethnic groups different from his own, (3) mastery of the use of words and numbers, (4) positive attitudes toward school and learning, (5) the habits and values associated with responsible citizenship, (6) good health habits and knowledge of the conditions necessary for maintaining physical and emotional well-being, (7) experience with creative processes in a variety of fields, (8) a full grasp of the opportunties for preparing for a productive life, (9) understanding and appreciation of the natural sciences, the social sciences, the humanities, and the arts, and (10) preparation for

---

[24] Antoinette T. Ryan, "Testing Instructional Approaches for Increased Learning," Phi Delta Kappa, 66:524-6, June, 1965.

[25] Paul R. Mort, W. S. Vincent, and C. A. Newell, The Growing Edge (New York: Metropolitan School Study Council, 1957).

[26] Hagen and Thorndike, loc. cit.

[27] Paul L. Dressel, "Measurement and Evaluation," Thirteenth Yearbook (1960) of American Association of College for Teacher Education (Washington: The Association, 1960), pp. 45-52.

effective participation in a changing world.[28]

These goals, of course, resemble others set forth by various groups in the past.

Basically, instruments and procedures used to measure objectives are limited only by the ingenuity of the evaluator and the availability of time and money. The use of tests of all types is well known; aptitudes, attitudes, and achievements are regularly measured with more or less validity. Even while old devices are revised toward improvement, new ones are being introduced. Webb, *et al.*, suggest an increased use of observation, review of archives (i. e., public records), and other techniques of which subjects would be unaware, to supplement the more traditional sources of information.[29]

Accreditation of schools, by use of self-evaluation procedures and visitation teams, is also a familiar process. School surveys are commonly used to evaluate special problem areas of school systems.

Appropriate research designs are obviously essential in evaluation. A helpful summary of several different types is given by Flanagan, who illustrates briefly with such examples as Project TALENT.[30]

The increasing availability of electronic data processing equipment and skills makes possible the collection and analysis of larger quantities of information than could have been effectively utilized in the past. Effectiveness of new programs, for example, can now more easily be evaluated by keeping track of individual students and their curricular experiences from a longitudinal point of view.[31]

If educators can remain "product-oriented" toward desirable goals, benefits in the form of improved evaluation may result from the introduction of new techniques developed primarily in the defense and aerospace industries and in the federal government. Systems analysis is now being utilized as an approach to both large- and small-scale educational problems.

---

[28] Dyer, *loc. cit.*

[29] Eugene J. Webb, *et al.*, *Unobtrusive Measures: Nonreactive Research in the Social Sciences* (Chicago: Rand McNally, 1966).

[30] John C. Flanagan, "Evaluating Educational Outcomes," *Science Education*, 50:248-51, April, 1966.

[31] Alvin Grossman and Robert L. Howe, *Data Processing for Educators* (Chicago: Educational Methods, Inc., 1965), p. 251.

Applications of PERT in the field of education are now being developed and used.[32] All of these techniques will probably be required if the demands of the newer concepts of program budgeting and cost-effectiveness (already in operation at the federal level) are to be met by local school systems.[33]

Conclusion

An effort has been made to provide sufficient background information to enable interested readers to pursue futher topics of their choice in the broad and dynamically expanding field of educational evaluation. Although evaluation is admittedly difficult, guideposts offered in this text and bibliography should be of some use to educators vitally interested in the evaluative process.

---

[32] Desmond L. Cook, An Introduction to PERT (Columbus, Ohio: Bureau of Educational Research and Services, Ohio State University, 1964).

[33] Werner Z. Hisch, "Education in the Program Budget," Program Budgeting (David Novick, ed.), (Washington: U. S. Government Printing Office, 1965), Chap. 6, pp. 131-54.

11

IMPLICATIONS OF VIEWING EDUCATIONAL EVALUATION AS RESEARCH IN THE BEHAVIORAL SCIENCES

Edwin P. Willems*

Under a flood of public financial support and pressure toward innovation, experimentation, and change, resulting in a concern about short-term and long-range effects of unusual support and change, evaluation has become a distinctive password in education circles.[1] As with many passwords, it is easy to hope for too much from evaluation, and to infuse it with magic and power that it does not have. There should be some advantage in stripping the concept of its unintended and undeserved attributes and preserving its more realistic and workable ones.

First, evaluation says nothing about what the goals of education ought to be. The generating and articulating of goals, hopes, and ideals come from community consensus and society's norms. Rather, evaluation is a process of gathering evidence that will form the basis for the most accurate statements possible about particular institutions, strategies, sequences, and styles of educational practice. While the resulting statements may focus on the extent to which goals and hopes are realized, evaluation itself is the process by which one uses the best skills and techniques available to make reasonable statements, that is, to answer questions about evaluation.

Second, evaluation as such is neutral as to the level at which the questions are asked and answers sought. At one time, the question might be, "To what extent does school system A meet certain standards of economic efficiency?" At another time, one might ask, "Is teaching style Z enhancing the reading speed of intellectually marginal students?" Futhermore, evaluation as such is neutral as to the time perspective of the questions asked. That is, one might wonder about a portion of a particular class session or about lifelong effects.

Viewing evaluation as the process by which persons use the best skills and means available to answer questions makes it equivalent to another set of human activities called research. Research has also become a password, and as a result, there are varying opinions about its power and usefulness. It is also common to infuse research with too much magic and power, and to circumscribe it with overly strict boundaries. For example, Kerlinger says, "It should be clear then, that the ideal of science (research)[2] is the controlled experiment...the controlled experiment is the desired model of science (research)."[3] No particular method or technique represents the model of research. The model of research is the process by which a person or group of persons use the best skills or means available to gather evidence to answer their questions. Thus, evaluation and research are similar, if not equivalent, processes.

---

*Assistant Professor of Psychology, Rice University.

[1] Gene Atkinson and David A. Knickel, "Evaluation of Educational Programs: An Exploration." (Paper prepared for Gulf Schools Supplementary Education Center, Pearland, Texas, 1966).

Assuming, as Kerlinger does, that one particular method or technique represents the model of evaluation or research has two subtle but pervasive by-products: (a) Every issue or question about education or behavior must be tailored to fit that method. (b) When the issues or questions do not fit that method, they are not amenable to evaluation or research. This is putting the cart before the horse. Tools and methods must be chosen by persons to suit their problems and questions.

Viewing evaluation as equivalent to, or as a special case of, research and viewing research as a personal, human process of seeking answers to questions, have many implications, some of which will be discussed below. However, one distinction must be made clearly at the outset, the distinction between "personal" and "private." Research is a personal activity, subject to many human foibles, mistakes, and weaknesses. However, research is also an explicit, documented process, open to public scrutiny, so that another person can independently repeat the process and independently verify or refute the resulting statements. A private, undocumented process for generating statements is a mysterious one that cannot be independently repeated and checked. Such a private process is not research. The issue here is the general issue of the grounds for evidence. Whenever someone makes statements in the evaluation of education, it should be possible to assess independently whether his statements are justified. If the process of arriving at his statements is private, the accuracy of his statements cannot be assessed.

Methodology in Behavioral Research

The argument can be taken a step futher by viewing evaluation as a special case of research in the social or behavioral sciences. That is, evaluation is research on institutions, organizations, large and small social groups, and individual abilities and predispositions, i.e., the interplay among physical, economic, and social environment and individual behavior and experience. The distinct advantage of viewing evaluation as social and behavioral research is that there exists an extensive literature on the issues, problems, strategies, techniques, and guidelines of such research. This literature is too far ranging to review here; however, authors such as Gage,[4] Kerlinger (mentioned above), Selltiz *et al.*,[5] Festinger and Katz,[6] and Webb *et al.*[7] have provided excellent treatments of research methodology. A great number of books and articles on educational and psychological testing are also available. In other words, the process of obtaining data to answer questions, together with the precautions and guidelines required for making statements based on the data, have been thoroughly described and documented. However, in spite of this extensive literature, there remains one special problem for education evaluation that has not been treated extensively enough and in enough detail. This problem has to do with what one is studying and the conditions under which one is studying it in the typical, workaday exercise of education evaluation.

---

[2] Present author's insertion.

[3] F. N. Kerlinger, Foundations of Behavioral Research: Educational and Psychological Inquiry (New York: Holt, Rinehart and Winston, 1964), p. 291.

Two Strategies of Research

One way to evaluate an educational program, or a part of an educational program, is to construct a paradigm case and study its outcomes, that is, to conduct an explicitly arranged experiment. This strategy has the great advantages of any controlled experiment, such as carefully isolating and controlling the conditions and it increases the possibility of making clear statements on what leads to what. Laboratory schools and experimental educational groups, in which students are carefully selected and assigned to groups and in which teaching conditions and environmental circumstances can be varied or held constant at will, are examples of this evaluative strategy. A futher advantage of this strategy, here called *experimental evaluation*, is that the rationale and guidelines for it have been discussed so carefully by writers such as Campbell and Stanley,[8]

---

[4] Nathaniel L. Gage, (ed.), *Handbook of Research on Teaching.* (Chicago: Rand McNally, 1963).

[5] Claire Selltiz, *et al.*, *Research Methods in Social Relations.* (New York: Holt, Rinehart and Winston, 1961).

[6] Leon Festinger and Daniel Katz, (eds.) *Research Methods in the Behavioral Sciences* (New York: Holt, Rinehart and Winston, 1953).

[7] Eugene J. Webb, *et al.*, *Unobtrusive Measures: Nonreactive Research in the Social Sciences* (Chicago: Rand McNally, 1966).

[8] Donald T. Campbell and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," *Handbook of Research on Teaching*, Nathaniel L. Gage, editor (Chicago: Rand McNally, 1963), pp. 171-246.

Shaw,[9] Festinger,[10] and Kerlinger (mentioned earlier), who offer strong and reasonable arguments favoring the experimental approach.

Granting the advantages and strengths of experimental evaluation, the major problem in adopting it as a strategy is that in the usual evaluation of an educational program, whether by a participant in the program or by an outside person or agency, that which is being evaluated is an already existing, intact program. By its very nature, the typical, workaday process of evaluation involves what is here called *naturalistic evaluation* or *naturalistic research* wherein the evaluator or researcher does not have manipulative, experimental control over the events he is studying.[11] In contrast to experimental research, there is a relative lack of careful, thoughtful literature that discusses the strengths, rationale, and guidelines for naturalistic research. In other words, despite a few exceptions,[12] there is a comparative lack of material dealing with the kind of research methods that the typical evaluative process calls for.

---

[9] M. E. Shaw, "Social Psychology and Group Processes," *Experimental Methods and Instrumentation in Psychology*, J. B. Sidowski, editor (New York: McGraw-Hill, 1966), pp. 607-643.

[10] Leon Festinger, "Laboratory Experiments," *Research Methods in the Behavioral Sciences*, Leon Festinger and Daniel Katz, editors (New York: Holt, Rinehart and Winston, 1953), pp. 607-643.

[11] The following paper discusses some of the differences between the experimental and naturalistic strategies:

Edwin P. Willems, "Toward an Explicit Rationale for Naturalistic Research Methods." *Human Development* (In press.).

[12] Roger G. Barker, "Observation of Behavior: Ecological Approaches," *Journal of Mt. Sinai Hospital*, 31:268-284, 1964;

Roger G. Barker, "Explorations in Ecological Psychology," *American Psychologist*, 20:1-14, 1965;

Paul V. Gump and Jacob S. Kounin, "Issues Raised by Ecological and Classical Research Efforts," *Merrill-Palmer Quarterly*, 6:145-152, 1959-1960;

Edwin P. Willems, "An Ecological Orientation in Psychology," *Merrill-Palmer Quarterly*, 11:317-343, 1965;

Edwin P. Willems, "Toward an Explicit Rationale for Naturalistic Research Methods," *Human Development* (In press.);

Edwin P. Willems, and H. L. Raush, (eds.) *Naturalistic Viewpoints in Psychological Research* (New York: Holt, Rinehart and Winston). (In preparation.).

Dimensions of Naturalistic Research

Because of the relative lack of detailed material on naturalistic research, some space will be devoted to the rationale and reasons for such methods.

Inventory of educational phenomena. Handbooks and encyclopedias of the physical sciences agree in saying that potassium ranks seventh in order of abundance of elements, and constitutes 2.59 percent of the igneous rocks of the earth's crust; that potassium compounds are widely distributed in the primary rocks, the oceans, the soil, plants, and animals; and that soluble potassium salts are present in all fertile soils. Such findings on the distribution of potassium in nature are extremely helpful to persons who are evaluating occurrences of potassium and its compounds and effects. In other words, physical scientists and engineers have detailed information on the frequency and range of occurrence of the things with which they deal. It is odd that apart from the occasional, spotty educational census, the norms of a few tests, and some selective service summaries, the engineers, planners, and evaluators of educational programs and behavior know so little about how their phenomena are distributed in nature. They know relatively little about the frequency, range, and classification of teaching styles, problem-solving, succeeding, failing, aspiring, cooperating, accepting responsibility, playing, being valued, conflict, etc. In other words, in the evaluation of programs for optimizing and arranging human resources, there is no handbook, inventory, or periodic table which will put specific findings into perspective. The process of generating such a handbook of education-related behavior calls for a tremendous amount of naturalistic research or the observation and classification of intact, naturally-occurring educational situations. Evaluation, if viewed and conducted as serious research, could aid in filling this need.

Problem of yield. It is often thought that naturalistic research, the kind of research involved in evaluating existing programs, is inefficient in that a disproportionately small amount of information is gleaned from the investment of time and money. Experimental evaluation is often viewed as being more efficient, in that the information yield is greater.

The problem of ratio of effort to yield occurred in a study by Gump and Kounin,[13] who had been studying how disciplinary techniques directed at one child affected the other children in classrooms. Extending their studies to public camps, they observed camp leaders and campers. In the written narratives describing 46 hours of the behavior of campers, only one incident of the type they were studying occurred for every 18 pages of narrative. This was extremely low yield from a naturalistic study. However, Kounin reported another study in which the yield was disappointingly low from a piece of experimental evaluation.[14] Kounin was again assessing effects of disciplinary techniques upon nontarget children, but this time he wanted to assess their affects upon children who ranged from high to low on commitment to the tasks. After carefully training his experimental teachers, he arranged for a sample of adolescents to come to the laboratory school during the summer, for pay, to serve as experimental students, while in intact, naturally-occuring classrooms, he was able to find such students. The implication is that the problem of economy of effort is not unique to naturalistic evaluation.

---

[13] Paul V. Gump and Jacob S. Kounin, "Issues Raised by Ecological and Classical Research Efforts," Merrill-Palmer Quarterly, 6:145-152, 1959-1960.

Artificiality. Perhaps more important than the question of yield as the liklihood of artificiality in experimental evaluation. Artificiality refers to educational arrangements and situations that students and teachers never confront in their intact, everyday experiences. Two published studies illustrate the problem of artificiality.

Gump and Kounin (mentioned above) wanted again to assess the effects of disciplinary techniques by teachers upon nontarget students, this time in college classrooms. They arranged for student accomplices to come late to classes, and they trained teachers to sometimes try supportive, friendly reactions and sometimes threatening, punitive reactions on the late arrivals. In this experimental arrangement, they found that punitive, as compared to friendly, reactions resulted in lower ratings of the teachers' competence, likableness, and fairness by nontarget students. However, the investigators found through later questionnaires that the students were surprised that one of their college teachers would make a point of taking time out from class to correct and respond to a late arrival. In their attempts to create a neat experimental evaluation, Gump and Kounin had taken a phenomenon out of its everyday context and produced an arrangement that was not customary for the students.

Gump and Sutton-Smith studied the reactions and behavior of poorly skilled boys when they were assigned to more or less difficult roles in games.[15] In the experimental game, a boy who was It worked in the center of a rectangular playing field and attempted to tag other players who ran to and from safe areas at each end of the field. High power was given to the It role by having the boy who was It call the turns; he dictated when others could run. In the low power condition, the other players could run whenever they chose. Preselected slow runners were assigned to the two It conditions. In the low power as compared to the high power It conditions, poorly skilled boys experienced more tagging failures, uttered more comments indicating defeat and distress, and were teased and combined against more frequently. However, when the investigators adopted the naturalistic strategy and observed what happened in gyms, camps, and playgrounds, where they had not arranged and manipulated the same conditions, they found that poorly skilled boys seldom got into difficult games. If they did get involved, they avoided or were exluded from the difficult roles. The experimental evaluation created conditions that were artificial and unusual for the boys. The naturalistic evaluations pointed out in what ways the experiment failed to match the everyday event.

---

[14] Jacob S. Kounin, "Dimensions of Adult-Child Relationships in the Classroom." Paper read at Topology Meeting, New York, August, 1961.

[15] Paul V. Gump and B. Sutton-Smith, "The It Role in Children's Games," The Group, 17:3-8, 1955.

There are several implications of these examples of artificiality. There is certainly nothing inherently wrong with experimental evaluations. In fact, experiments are often ideal for testing hunches and theories. On the other hand, it is clear that naturalistic evaluation is not a weak second cousin of experimental evaluation. In fact, it is the only choice of strategies when one wishes to evaluate intact, already existing educational arrangements. Finally, as a later section will suggest, naturalistic evaluation can often point out specifically those ways in which experimental educational research has generated artificial, restricted conditions.

Repeatability. Reference was made earlier to a distinction between personal and private activities, and the point was made that private activities have no place in research or, therefore, in evaluation. One characteristic of evaluation that takes it out of the private domain and brings it into the acceptable research domain is repeatability. An evaluation is repeatable when the same evaluator or another evaluator can use the same techniques or observe the same conditions a second or third time and end up with the same results. If two persons evaluate the same phenomenon, supposedly using the same techniques of observation and assessment, and get discrepant results, the evaluative processes they have used should be open to question as containing too many private, unrepeatable aspects. Naturalistic research, the type typically required for evaluating existing educational conditions, is often criticized on the grounds that it is not repeatable. Fortunately, repeatability is not a matter for pure speculation and argument to decide; it can be decided on the basis of what persons actually find. One example will illustrate the point.

In the face of a strong movement to consolidate high schools in Kansas, a group of psychologists sought to evaluate the effects of school size upon the experience and behavior of students.[16] One subpart of this program of research, which is still in progress, sought to assess how school size related to the development of sense of obligation to school activities, or social responsibility, in the students. Two studies, four years apart, involving different samples of students, yielded almost identical results: students in small schools reported much more sense of social obligation than students in large schools. Several studies independently obtained corroborating results. The point here is that all of these studies involved naturalistic methods of evaluation, and they yielded repeatable, nonprivate results.

---

[16] Roger B. Barker and Paul V. Gump, Big School, Small School: High School Size and Student Behavior (Stanford: Stanford University Press, 1964).

Generalizing. Much of the preceding discussion is relevant to what is perhaps the most important rationale for naturalistic evaluation, or naturalistic research. It is probably a truism that persons who are evaluating an educational arrangement wish to make general concluding statements about the arrangement and perhaps others like it. Several examples will illustrate how naturalistic research sometimes disagrees with experimental research and will suggest how naturalistic evaluation serves a uniquely important function. The examples may appear to range far afield from direct educational applications, but their implications are rather direct and important.

S. L. Washburn, the anthropologist, reports that baboons have been studied frequently in experimental research, as in primate laboratories, and in captivity, as in zoos.[17] When two or more baboons are confined together, it is common to find that one emerges as the leader, which is not surprising in itself. Under these conditions, the leaders emerge and maintain their leadership through physical intimidation and comparative brute power. Washburn goes on to report the results of observations of baboons in their natural habitats, and one set of findings is of special interest. Leaders also emerge in the wild, but they appear not to emerge by physical intimidation and brute power. Superior cunning, sexual expertise, and attractiveness seem to be the route to leadership among baboons in the wild.

The work of Beecher on pain relieving drugs also suggest the important contribution of naturalistic research.[18] Beecher's findings suggest

---

[17] Sherwood L. Washburn, Phi Beta Kappa Lecture, University of Kansas, 1963.

[18] The following references summarize some of Beecher's research:

Henry K. Beecher, "Relationship of Significance of Wound to Pain Experienced," Journal of the American Medical Association, 161:1609-1613, 1956;

Henry K. Beecher, "Generalization From Pain of Various Types and that laboratory experiments on pain and pain-relieving drugs contribute little to the understanding of pain and analgesia as they occur outside the laboratory. Compounds that seem to relieve pain in the controlled laboratory conditions often fail to do so in the wards of hospitals, while substances which, by laboratory experiment would seem to be ineffective frequently are high effective in the clinical situation. In other words, the picture of drugs and pain relief would be incomplete without naturalistic evaluation.

Hall and Williams, two psychologists, recently reported a fascinating study of how groups handle conflict and solve problems.[19] They compared the performances of 20 established, intact groups of management trainees to 20 contrived groups, groups with no history as groups before the individuals came to the laboratory. The established groups differed markedly from the contrived groups, expecially in the way they dealt with conflict. In other words, Hall and Williams found that it is precarious business to generalize from contrived laboratory groups to established, naturally occurring groups.

The implications of these examples for educational evaluation are clear. It is seductively easy to be content with the results obtained in controlled experimental research and to assume that caution is the only requirement for generalizing these findings and making them applicable to the real life situation. The examples suggest that caution is not the only requirement, and that naturalistic research performs a uniquely required function. Not only is naturalistic research not a second-rate adjunct to experimental research, but it is very important in ascertaining the generalizability of experimental research. In other words, naturalistic evaluation promises to perform an important function in educational research.

Summary of naturalistic evaluation. A significant amount of space has been devoted to the rationale for naturalistic evaluation, and parts of the discussion appear propagandistic. This is so by intent, for two reasons. First, evaluation is a special case of research. The kind of

---

Origins," *Science*, 130:267-268, 1959;

Henry K. Beecher, *Measurement of Subjective Responses* (New York: Oxford, 1959);

Henry K. Beecher, "Increased Stress and Effectiveness of Placebos and Active Drugs, " *Science*, 132:91-92, 1960.

[19] Jay Hall and Martha S. Williams, "A Comparison of Decision-making Performances in Established and *ad hoc* Groups," *Journal of Personality and Social Psychology*, 3:214-222, 1966.

research represented by the evaluation of intact, already existing situations can and should be recognized by all those who take educational research seriously.

Advantages and Implications

In behavioral research, the investigator must have a clear idea of what he wants to study. In evaluation, one must have a correspondingly clear idea of what one is evaluating. In both research and evaluation, the investigator seeks to make statements that document the occurrence of something or that document that one thing leads to another. In all these efforts, the investigator or evaluator attempts to gather information or data to form a basis for his statements. One advantage of viewing evaluation as a special case of research is that other persons engaged in research have learned certain critical lessons from the process of gathering data. These lessons have been well-documented, and should also be taken into account in evaluation.

Reliability. One such important lesson concerns the reliability of the ways, means, and techniques used to gather information. Tests, first-hand observations, questionnaires, interviews, tabulations from school records, and reports of students and teachers are all examples of techniques that might be used for gathering evaluative information, and they constitute what researchers call measures, or data-gathering techniques. If one administered an achievement test twice to the same group of students, under standard conditions and found that students' scores fluctuated widely from one time to the next, one should question the ability of the test to measure anything. This is one example of lack of reliability. Authors of published tests usually study reliability as a matter of course. However, it is easy to forget that any technique for gathering information is a measuring instrument whose reliability should be assessed directly. If a technique or measuring instrument fails the test of reliability, then the answer to the critical question, "Does my technique measure anything?" must be case in doubt.

The problem of realiability is not critical in measures of such things as weight, height, hair color, and sex, but measures of the kinds of things educational evaluation seeks to assess usually involve very complicated problems of reliability. If a zoologist has difficulty arriving at reliable measures of how many chimpanzees cross paths in a small jungle area,[20] it seems likely that evaluation of education will have difficulty arriving at reliable measures of much phenomena as self-actualization or "a full grasp of the opportunties for preparing for a productive life."[21] Futhermore, it will not do for persons involved in evaluation to deemphasize and criticize those who worry about reliability, because reliability is fundamentally involved in what constitutes acceptable evaluative evidence.

---

[20] Adriaan Kortlandt, "Chimpanzees in the Wild," Scientific American, 206:128-140, May 1962.

Validity. Reliability focuses on the question, "Do I have a measure at all?" Assuming reliability, the next critical question is, "Does my technique measure what I *think* it measures?" The latter is a question of validity, a question that has been thoroughly probed in the literature on research.[22] By way of analogy, suppose one were informed that school A had beaten school B by a score of 14 to 10 in football. The scores would undoubtedly represent valid measures of the number of times the teams scored touchdowns, extra points, and field goals; however, they would probably represent increasingly *invalid* measures of team ability, quality of coaching, and quality of football in the regions from which the schools came. Likewise, data on how often students are absent are probably a valid measure of attendance, but their validity as a measure of student initiative or sense of responsibility is much more questionable.

Validity is directly involved in the grounds or basis for making evaluative statements, and the question, "Do his data measure what he thinks they measure?" probably lies at the heart of many controversies that arise over evaluation.

Control groups. Closely related to the problem of validity is the importance of control groups, another lesson learned the hard way by persons involved in research. Campbell and Stanley, in the paper cited earlier, discuss the usefulness and importance of control groups in detail, so that only enough space will be devoted to them here to suggest their relevance to evaluation. One detailed example will suffice.

Sense of responsibility, commitment, loyalty, or sense of obligation, all refer to something that parents, teachers, and group leaders commonly hope children will acquire during their development. Several years ago some studies were conducted to assess the relation between high school size and students' sense of obligation to school activities.[23] For present purposes, the exact details are less important than the findings. Students in small high schools reported a strong sense of obligation to activities, much stronger than their counterparts in a large school. However, the small schools were located in small rural towns, while the large school was located in an urban area, a circumstance that made it impossible to attribute the differences in sense of obligation to school size instead of rural-urban differences. In this case, another study served the function of a control group.[24] The second study compared students in schools that varied in size but which were all located in relatively small towns, and obtained the same results for school size. Taking into account this control study, the impact of school size upon sense of obligation could be evaluated more clearly. In other words, rather than controlling anything in the usual sense, control groups, when properly chosen and used, allow one to attribute effects where they should be attributed, a very important consideration in educational evaluation.

---

[21] Henry S. Dyer, "The Pennsylvania Plan," *Science Education*, 50:242-248, 1966.

[22] Validity is one of the most common topics in the literature on educational, psychological, and social research. The following is just one example of such discussions:

Lee J. Conbach, *Essentials of Psychological Testing*. 2nd ed. (New York: Harper and Row, 1960).

Only three implications, in the form of lessons to be learned, that follow from viewing evaluation as a special case of research have been discussed in detail. The three areas discussed, reliability, validity, and control groups, deal with the process of gathering defensible information, and are perhaps the most fundamental. Other lessons learned from research also have their implications, and will only be mentioned. Here such specific evaluative problems as construction and use of questionnaires and tests, proper sampling of circumstances and persons, interview techniques, and useful, informative statistical techniques for summarizing and presenting the information gathered in evaluation all have their own unique problems and have all been treated extensively in the literature.[25]

---

[23] The following papers summarize the research on sense of obligation:

Edwin P. Willems, "Forces Toward Participation in Behavior Settings." In Roger G. Barker and Paul V. Gump, Big School, Small School: High School Size and Student Behavior (Stanford: Stanford University Press, 1964), pp. 115-135;

Edwin P. Willems, "Participation in Behavior Settings in Relation to Three Variables: Size of Behavior Settings, Marginality of Persons, and Sensitivity to Audiences." Unpublished doctoral dissertation, University of Kansas, 1965;

Edwin P. Willems, "Sense of Obligation to School Activities in Relation to Size of School and Type of Student," Child Development (Paper submitted and under review).

[24] W. J. Campbell, "Some Effects of High School Consolidation." In Roger G. Barker and Paul V. Gump, Big School, Small School: High School Size and Student Behavior (Stanford: Stanford University Press, 1964), pp. 139-153.

Simplifying in Evaluation

The reasons for, and advantages and implications of, viewing evaluation as a special case of behavioral research discussed so far focus on factors that might usefully be taken into account in the actual exercise and activity of evaluating. One of the earmarks of research and, therefore, of evaluation, is that evaluative problems be refined, narrowed, and simplified into manageable size so that reasonable, defensible information can be gathered. No one to date has satisfactorily evaluated education per se; every attempt at evaluation selects some aspect or aspects of the educational process as being of special interest or particular importance. Sometimes, insidiously, the selection is sometimes made according to the kinds of observational and measuring tools that happen to be available. No matter what the reasons for the selection, the typical approach to educational evaluation is a simplifying one, where simplifying means arbitrary untying, dismantling, and selection from the complex phenomena of education. The arguments in favor of simplifying are compelling, but simplifying also has major drawbacks. Just one such drawback, common to research in general and to evaluation in particular, is the very strong human tendency toward faddishness. In behavioral research and in evaluation it is tempting to invest the piece of work one is doing with too much value and representativeness, and assume that what one is observing and studying represents, or is, education. In the overall effort called evaluation, not only is there room for, but there should be explicit attention given to widening, complicating, and enriching the scope and complexity of the set of things and occurrences to be evaluated. In other words, there is also a need for just the opposite of simplifying, for at least three reasons.

The first such reason is the common tendency to develop stereotypes about what education is. One such stereotype might view education as the circumstances that produce creative problem-solving in students. Stereotypes have the advantage of leading to crucial, clear-cut evaluations, for example, measures of creative problem-solving on students from different schools and classrooms. However, stereotypes can also be translated into exclusive educational goals, whereby, for example, one might assume that if a little creative problem-solving is good and useful, more should be better.

---

[25] See, for example, Handbook of Research on Teaching, edited by Gage and mentioned above, and various chapters in:

Gardner Lindze, (ed.) Handbook of Social Psychology (Reading, Massachusetts: Addison-Wesley, 1954).

In addition to the tendency to stereotype education itself, a second reason for guarding against premature simplifying is a common tendency in research, called "the law of the hammer."[26] The law of the hammer goes as follows: If you give a child a hammer, things to be pounded become the most important things around. Likewise, it is too often true that if one gives the researcher or educational evaluator a standardized achievement test, achievement _as that test measures it_ becomes the most important aspect of education to be evaluated. Of course, the same can happen when the evaluator has a pet way of measuring self-actualization, leadership, social responsibility, ego-strength, student satisfaction, problem-solving, and so on, but the tendency encourages premature simplifying.

The third reason for avoiding premature simplifying focuses on the admirable but misleading little phrase: "Let's look at the facts." The directive to "look at the facts" can often cut away much of the polemic and controversy in educational evaluation, but the response to it can also be subtly misdirected. As a case in point, the Department of Education of the State of Kansas published a comprehensive educational survey of Kansas schools in the late 1950's, and the survey itself stated that any evaluation of the recommendations of the survey, without knowledge of the facts on which they were based, would be of little value. From the sizes, curricula, programs, and administrative organizations of Kansas schools, the survey concluded that the schools were suffering from educational malnutrition. Unfortunately, the survey did _not_ provide the factual basis for a diagnosis of malnutrition, which should have included careful examination of the children who partook of the fare provided. To provide the factual basis, a group of psychologists at the University of Kansas undertook a series of studies of Kansas schools (see Barker and Gump, mentioned earlier), and found that when the experience and performance of the students were considered, some of the survey's conclusions were unfounded. These psychologists took seriously the dictum, "Let's look at the facts," indicated the ways in which the picture painted by the survey was prematurely simplified, and strengthened the picture by complicating it.

A second case in point is the history of attempts in psychology to evaluate the effectiveness of various types of psychotherapy. In the early 1950's, the cry went up, "Let's just look at the facts," and various types of psychotherapy were evaluated in terms of simple discharge rates.[27]

---

[26] From a personal conversation with Professor Egon Guba.

[27] H. J. Eysenck, "The Effects of Psychotherapy: An Evaluation," _Journal of Consulting Psychology_, 16:319-324, 1966.

These early attempts were completely inconclusive, and it is only now, fourteen years later, that the complexity of evaluating psychotherapy and guidelines for evaluating it are taking useful shape.[28]

All evaluation requires some simplification and choice. However, keeping in mind (a) the common tendency to enshroud the concept of education in stereotypes, (b) the common tendency to define education by the available measuring tools, and (c) how misleading the call to "Look at the facts" can be, might prevent premature and overly narrow simplification.

Evaluation and Everyday Practice

All that has been said so far has been directed toward the implications of viewing evaluation as research and the special contributions that evaluation can make to educational research in general. Little has been said about the relation between evaluation, or the persons engaged in evaluation, and the actual, day-to-day practice of education in the hands of school boards, administrators, teachers, and students. In the eye of the day-to-day practitioner evaluation, especially when it is done by third persons and outsiders, is often seen as nosiness, intrusion, criticism, and recommended change. Perhaps, viewing evaluation as research will intensify such feelings. However, advocating a view of evaluation as research in no way implies a position that is incompatible with day-to-day practices, commitments, and traditional wisdom.

If over the years many different approaches of the practice of education have been tried, if some approaches have worked better than others, and if those that have worked better have, to some extent, been more persistently practiced by their originators, or imitated by others, or taught to apprentices, then the practices that have emerged may represent a valuable and tested portion of all possible practices. However, the bases for selecting and pruning among practices are very imprecise in the day-to-day process. Sometimes, what survives is a product of chance, or whim, or habit. Explicit evaluation, perhaps by an outsider, enters here as a means for sharpening the selection process. Explicit evaluation is not a source of ideas that are necessarily contradictory with day-to-day, accumulated wisdom, but is rather a refining process, or an aid to the refining process, supplementary to and sometimes superimposed upon accumulated wisdom. Thus, everyday accumulated practice and evaluation are both products of accumulation, remaining from experiences from which much has been weeded out. Evaluation as viewed here is one source of evidence for making decisions, and is not itself a policy-planning activity.

---

[28] D. J. Kiesler, "Some Myths of Psychotherapy Research and the Search for a Paradigm," Psychological Bulletin, 65:110-136, 1966.

III

SUBCULTURES OF THE SCHOOL; SELF-ACTUALIZATION AND STUDENT EVALUATION

Charles L. Mulford*

Two developments have been noted in recent social-psychological considerations of the school. First, it is widely recognized that it is useful to view the school in terms of its unique culture or ethos. As noted by Kallenbach and Hodges,[1] "There is no doubt that each school has a culture of its own. These cultures can range from the comparatively homogeneous rural ..., to the diverse patterns of an urban secondary school of three thousand or more students." The relationship between the school's culture, student behavior, and effective evaluation techniques will be discussed in this paper.

Secondly, recent research has suggested that the appropriate unit of analysis should be "the-student-in-the-school." By this it is meant that the social needs and satisfactions of students should be considered vis-a-vis the requirements of the school as an efficient organization (having a formal organization structure, directive leadership, and aspects of managerial control). These researchers specifically suggest that the social-psychological needs of students may not be congruent with some organizational needs and requirements. Student "self-actualization"[2] is a term used to indicate the degree to which students feel that their personal needs are being met and/or considered by the organization (school) officials. A discussion of the impact of the culture of the school on student behavior will be followed by a discussion of the dynamics of student self-actualization.

Why is a knowledge of the school's culture(s) or ethos necessary? Becker[3] suggests that "culture (or reality) shock" may occur if the teacher's expectations of his pupils' behavior do not "fit" closely to what actually occurs. For example, "teachers find that some of their pupils act in such a way as to make themselves unacceptable in terms of the moral values centered around health and cleanliness, sex and aggression, ambition and work, and relation of age groups." It is not necessary for the teacher to adopt the attitudes and patterns of behavior of his pupils, but foreknowledge of these (student cultures) will facilitate a person's role as educational leader.

---

*Assistant Professor of Sociology, University of Houston.

[1] W. Warren Kallenback and Harold M. Hodges, Sr., Education and Society (Columbus, Ohio: Charles E. Merrill Book, Inc., 1963).

[2] Charles L. Mulford, "Self-Actualization in a College Environment," Journal of College Student Personnel (To be printed March 1967).

[3] Howard S. Becker, "Social-Class Variations in the Teacher-Pupil Relationship," Education and Society. Edited by W. Warren Kallenbach and Harold M. Hodges, Jr. (Columbus, Ohio: Charles E. Merrill Books, Inc., 1963).

Specifically, how does a student's cultural "heritage" and value orientations relate to his school behavior? First, consider the influence of the student's social class background. Hollingshead[4] reported in his study of "Elmstown Youth" that: the lower the social class of the students' family (1) the higher the drop-out rate, (2) the lower the vocational aspiration, (3) the lower the proportion planning on college, (4) the lower the course grades and the higher the failures, (5) the higher the proportion in commercial and vocational curricula, (6) the lower the education of the parents, (7) the lower the intelligence scores, and (8) the greater the number of recorded discipline problems. Wilson[5] adds, "Social class differences in educability stem from family differences in stimulation, training, and motivation." The home of the lower-class child, he suggests, is (I would temper this with "may be") no launching pad for upper-class orbiting. But, as Wilson further notes, upper and middle class families sometimes fail to reproduce themselves, both culturally and biologically. So, the schools must be able to identify talented students from all social classes, train them in the necessary skills, values, and attitudes, and send them into responsible positions in the community.

In summary, a student's cultural heritage (often different from his teacher's) preconditions him with regard to attitudes, values, perspectives, and knowledge. A teacher's understanding of the pervasiveness of social class influences should help to better understand the needs and potentials of his students. For example, a teacher should be able to answer these questions:

1. To which social classes do students in his classroom belong?
2. How does a student's social class influence his preparation for and selection of life goals?
3. When a teacher's social class heritage is different from that of his students, does this present problems? How are these problems typically resolved?

---

[4] Discussed in Everett K. Wilson, Sociology (Homewood, Illinois: The Dorsey Press, 1966).

[5] *Ibid.*

4. Can teachers identify students who are upwardly mobile, that is, those most likely to be selected for social status positions higher than those of their parents? What seems to be their special needs?

5. To what extent do teachers realize that students' behavior, achievement, and success may in part be a function of their social class heritage?

Not only does the student's cultural heritage affect his behavior, but in addition, he may also be affected by the dominant values and practices among his classmates. Coleman[6] suggests that the mediocre student rather than the talented is often rewarded. He found that the grades of "A" go to the truly exceptional student only where the *student population* itself honors academic achievement. Where the intellectual pursuits are not esteemed in the student subculture, students of lesser ability receive high grade records. The more talented students may tend to seek social status and recognition from their peers, whose value orientation they share, in athletics and social popularity. In this same study, Coleman notes that these social rewards have very little to do with facilities and equipment.

But observation tells any experienced teacher that not one but several subcultures or systems of student attitudes and values may exist in a school. Newly arrived students may encounter a cultural environment often diverse in its manifestations. Clark and Trow suggest that four subcultures typically are present: the academic, collegiate, nonconformist, and vocational subcultures. Clark and Trow's typology of possible student subcultures is shown below.[7]

Identify with School		Involved with Ideas
	Yes	No
Yes	Academic	Collegiate
No	Nonconformist	Vocational

The collegiate subculture is the world of dates, sports, and fun opportunities that are possible within and concomitant with school. The student who identifies with the school because of these socially rewarding opportunities but does not identify with the world of ideas is said to be a member of the collegiate subculture. A second orientation is the vocational subculture. In this subworld, students are narrowly concerned with job opportunities and preparations. A school pursuit is valued only if it will enable the student to "get a job." "Like the collegiate, the vocational subculture is resistant to intellectual demands beyond what is required to pass the course."[8]

---

[6] James S. Coleman, "The Adolescent Subculture and Academic Achievement," *American Journal of Sociology*, (Jan. 1960), p. 337 *et passim*.

[7] The author is leaning quite heavily in this section on an adaptation of Burton R. Clark's and Martin A. Trow's "Determinants of College Student Subculture" presented in *Sociology* by Leonard Broom and Phillip Selznick, (New York: Harper and Row, 1958), 3rd ed. pp. 453-456.

A third group of students identify with the academic subculture. That is, they identify with both the school and world of ideas. These serious students identify with the intellectual concerns of the teachers but may be regarded as "grinds" by some students.

A fourth orientation is the nonconformist group. "In the eyes of their more conforming classmates they are the unwashed... These somewhat rebellious students are concerned with the intellectual pursuits but not with the school as an institution. It [the subculture] offers the rebellious student shelter and intellectual support for his rebellious idealism."[9]

It is generally assumed that the relative importance of the collegiate subculture has declined in recent years, during which in the same period of time there has been a relative "triumph of vocationalism." However, a recent survey[10] suggested that only 19 percent of entering college freshmen from 23 different schools regarded themselves as "academic", 4 percent described themselves as "nonconformist", 27 percent as "vocational", and 51 percent described themselves as "collegiate" (fun and games).

To what extent should a teacher be able to answer these questions:

1. What subcultures exist in his school? How many students belong to each subculture?

2. Which subculture seems to be dominant?

3. How do teachers deal with members of the different subcultures; are teaching and counseling primarily oriented towards the needs of only one subculture? Which?

---

[8] _Ibid._

[9] _Ibid._

[10] Student Life Division, _Faculty and Staff Bulletin, No. 3_ (Houston, Texas: University of Houston), Nov. 9, 1966.

It was noted earlier in this paper that a second major concern has focused on whether or not there may be a lack of congruency between the social-psychological needs of students (student self-actualization) and the organizational requirements of the educational system. It is recognized, of course, that any individual involved in any relationship (including students enrolled in schools) necessarily surrenders some "freedoms" and degrees of personal autonomy. A healthy balance between social needs and organization requirements is necessary. Research[11] indicates that student self-actualization may be significantly related to school academic success and achievement. In this study of college students, a test was used that has been developed to measure student self-actualization. Scores from this test, which were significantly associated with academic success, were relatively independent from other traditional variables used in prediction studies such as sex, social class, high school preparation, and so on.

Are these questions relevant for teachers:

1. What are the important predispositions (needs and desires "expected" to be met) of students in this school?

2. What attitudes do teachers have toward these perceived needs and desires?

3. How can teachers structure their classroom so that maximum self-actualization will result?

It may be that the two concepts, subculture identification and self-actualization, can be profitably considered as part and parcel of the same problem. Put the question this way: which students are most likely to have their needs met in the classroom? That is, toward which subculture do most teachers direct their best talent and resources?

An almost impossible dilemma seems apparent. Unless all students come to identify with the teacher's orientation, a most unlikely event in most large urban schools, the only alternative may be that teachers will have to realize that different groups of students have different needs. Quoting from the recent survey of college students mentioned above, "The figures cited not only bear out the thesis that a minority of students can be described as primarily academic in orientation, they indicate also the strong concern of other students with the emotional aspects of their lives. Love and affection, self-respect, and personal identity seem to be in the center of their concern."[12]

---

[11] Mulford, loc. cit.

[12] Student Life Division, loc. cit.

In summary, each student (and each teacher) brings with him a cultural heritage stemming largely from his social class membership. A teacher's understanding of the many implications of social class may help him better meet the needs of his students. In addition to this influence, students may relate to another (independent of social class influences) cultural reality--one of several possible student subculture orientations, each with a different theme of life or ethos.

To maximize mental health, self-actualization, and objective achievement for each student is certainly a worthy goal. If this goal is to be met, however, a teacher's teaching, counselling, advice, and evaluation of his students may have to be extremely varied and rich in its content and form. These issues call for additional study and research so that the teacher may be aided in this endeavor.

IV

EVALUATION OF SCHOOL CONTROL RELATIONSHIPS

Bancroft C. Henderson*

The act of evaluation is the verbal or symbolic human process of applying values to some aspect of a society. Values are the degree of preference that the evaluator applies to a facet of a society. All phenomena for any one time or place. There are conflicting values, unconscious values, flexible changing values, ambiguous values, and values associated with roles. Values are relative and not absolute; they change with time, place, and role. They differ between societies. Within a society they differ in relationship to subsystems like the schools or between roles such as superintendent and teacher.

The collective act of administering evaluation, rather than individual evaluation, could serve the purposes of a school system by focusing attention on alternatives, by enriching the dialogue concerning school policies and their consequence, by increasing the awareness of values implicit in all policy, and by making more conscious the needs for research in relationship to policies and flexibility in regard to those policies.

Evaluation of control relationships places an emphasis upon the dominant-submissive relationships which are customary in bureaucracies with hierarchical structures. Control relations are then those aspects of a role which define formally, or informally, the discretion of the role partners, the degree of obedience or compliance which is expected, and the amount of initiation for change. Control is related to written regulation, hiring and firing authority, promotion and demotion authority, and other forms of sanctions. The relationship is important; it deals with a series of important questions. Who makes final policy? How independent should professional people be of direct supervision by nonprofessionals? Are teachers professional persons? How much responsibility should be left with the student and teacher in order to encourage mature self-regulation? How much should a school system cut itself off from the student subculture? Can democratic values be learned in an authoritarian setting? Should the public run our schools? How should the public participate in school policy formation? These and other critical questions are related to school control and its evaluation.

---

*Associate Professor of Political Science, University of Houston.

Controls from Outside the School System

If the system is defined as the school board and all of the personnel, students, and property of the schools per se, then the outside controls come from units of government and from public elections.

Since outside control factors are also outside the policy discretion of the school system, the type of questions considered in evaluation should not be entirely focused upon the wisdom of the policy itself, but should emphasize the type and style of compliance. For the purpose of evaluation of schools this is significant because the list of limits or controls is long and dramatic. It includes from many states:[1]

1. Keeping records of student attendance
2. Making general reports on students
3. Checking sight and hearing of students
4. Making inventories of school property
5. Using state adopted textbooks
6. Reading Bible daily without comment
7. Following established course of study
8. Supervising fire drills
9. Teaching health education with reference to use of alcohol and narcotics
10. Teaching certain ideas basic to good citizenship
11. Requiring loyalty oath by teachers
12. Refusing teachers right to wear religious garb
13. Requiring teachers to attend conferences or institutes
14. Requiring teachers to contribute to retirement fund
15. Requiring teachers to undergo annual physical health examination
16. Demanding teachers have no obligation to book companies

When you add this list to the limits set by taking restraints and financial aid as well as certification requirements for individuals and accreditation requirements for systems, the authority outside the system seems great. However, there are grants of authority to the levels within the system that are also extensive. These will be considered at appropriate points later in this paper.

Evaluation of controls at this point can involve some significant questions. How well aware are participants of these controls? Is there a handbook that describes the rules and regulations? How much support do these limits get from boards, administrators, and students? Do associations of boards, supervisors, and teachers express any opinion or have any voice in the writing of such policies? Are individual persons encouraged to participate in the formation of these policies? Are the policies obeyed? Are students aware of any noncompliance? These questions, as those included later, are of a "fact" nature and are intended as a basis for stimulating evaluation. When these kinds of questions are developed into an evaluation format they might cause discussions of significant issues.

---

[1] Leo Martin Chamberlin, The Teacher and School Organization (New York: Prentice-Hall, Inc., 1958), pp. 64-66.

The appropriateness of full participation of professional teachers in policies that regulate the profession is one of the most significant issues. The doctor dominates policy regarding doctors as do engineers and lawyers for their professions. This does not seem to be the case for teachers.

In another respect there are some significant control issues that stand as examples of the heart of the American political and social system. The way a system reacts to these issues will demonstrate to the student more significantly than any lecture the commitment to values related to these issues. Examples of such issues are school racial segregation, public aid to church schools, released time for religious instruction, religious prayers, flag salute requirements, teacher loyalty oaths, and legislative investigation of the beliefs of teachers.

If a school system vigorously participates in policies to reduce outside control, there could be in a general way an adverse reaction in the minds of students as the local school system would be vigorously protecting its "right of self-determination." The effect could be the impression of narrow provincialism plus hostility to government and politics. This may be an unfortunate message to inadvertantly communicate to students and the public. This possibility should be evaluated.

Since politics is conflict and conflict resolution, participation in politics is also significant to the resolution of conflicts over educational policy. Such conflicts as vocation vs. academic emphasis, much sex education vs. little concern, foreign language offering vs. little attention, and even the issue of whether controversy shall be encouraged, tolerated, or penalized are vital. Thus while vigorous political participation which is aimed at both limiting and affecting outside influence may communicate one set of values, nonparticipation would demonstrate political apathy and would leave the resolution of educational conflict issues to others who need the conceptualization of alternatives that a professional could offer. If the ideal is "expert" participation with a basis in professonal recognition, then the evaluation instrument should focus upon the degree to which this is a practice.

School Board Control

There is a broad range of authority that a school board may exercise. Following are some categories:[2]

1. May regulate dress
2. May require physicals
3. May exclude those jeopardizing health of others
4. May require vaccination
5. Can expel persons who join fraternities or secret societies[3]
6. Can exclude married students from activities such as football[4]
7. May prescribe required curriculum
8. May prescribe teacher method
9. Can expel a student who breaks any rule
10. Can expel for misconduct that does not break a rule[5]
11. May expel a student without notice of hearing
    (court cases may have modified this point)

This list is included to indicate that authority is extensive. The limits upon the use of this authority are thus within the system. Since the authority over teachers and administrators is also extensive and most teachers do not have the right to effective resistance, this issue of board control is critical.[6] Evaluation may cover some of the following questions which could raise the issue of appropriateness in relationship to significant values, for what is legal may not be valuable.

Are hearings and appeals provided in all disciplinary cases involving teachers, students, and administrators? Is hiring, promoting, and salary on the basis of merit or influence? Are citizenship rights of students protected or is in _loco parentis_ the perspective? How well are principles of free speech protected for students and faculty? Are there multiple channels of communication to the board, or is the superintendent the only channel? Do boards or their members belong to board associations? Are board members well-informed? Do they attend meetings regularly? Is representation on the board fairly apportioned in the population? Are school board election contests based on significant issues? Are the issues fully discussed and understood by the public? Is the board primarily a policy-maker and not administrator?[7]

---

[2] Lee O. Garber, _The Law Governing School Board Members and School Board Meetings_ (Danville, Illinois: Interstate Printers and Publishers, 1963), pp. 7-8.

[3] 18 Ca. App. 19.

[4] 330 S. W. (2d) 70F (Tex.) 1955.

[5] 87 (Ark.) 254.

[6] Garber, _op. cit._, No. 3., p. 6.

Many of these questions deal not just with the nature of control but with the style of control and the degree that the rest of the school system is likely to respect and accept the controls of the board as being representative of public thinking and based upon adequate information. When policies are soundly based, teachers and administrators find it difficult to resist them. Some of these types of questions were indicated by research that has been conducted concerning superintendents' expectations of school boards.

If these types of issues are incorporated into an evaluation instrument, the effectiveness of the board in relation to the system would be indicated.

The Administrative Role and School Control

Since this paper is short and intended only to be indicative of evaluation concerns, this section will combine all administrative levels in order to remain at a general level.

The subject of administration and control is related to the existence of a profession of education administration. One characteristic of a profession is the degree of autonomy of the administrator and the respect that is shown for his expertise.

Donald E. Tope writes about the development of a profession of education administration.

> Such an establishment must assess the current status of the school administration, the standards of selection and credentialing, the quality of the preparation programs designed to equip him for the job, and the social conditions necessary for him to perform his task effectively.[8]

It would appear from this statement and related remarks that the primary social condition necessary would be a high degree of self regulation, selection, and training of the professional by the professional. The areas of professional administrative authority deserving evaluation could be:

---

[7] Keith Goldhammer, The School Board 1964 (New York Center for Applied Research).

[8] Donald E. Tope, The Social Science View of School Administration (Englewood, New Jersey: Prentice-Hall, 1965), p. 2.

1) Degree of autonomy of control and the degree of initiative allowed, for the establishment of preservice and inservice education for prospective administrators[9]

2) Degree of access to board policymakers

3) Freedom and encouragement for affiliation with professional associations

4) Establishment of administrative organizational pattern

5) Setting of job description qualifications.

The degree to which the appropriate administrative levels dominate these policy areas and processes would be a measure of professionalism. An evaluation instrument could appropriately include this topic.

Another topic related to administrative evaluation could be the degree to which the administrative levels have secured board support. There is some evidence that this could be a problem related to policy. As an example, school board members and superintendents were asked, "Is it desirable or undesirable to give numerical grades on regular report cards in the first six grades?" Fifty-three percent of school board members said that it was desirable while only 9 percent of superintendents agreed.[10] This would indicate that there may not have been an informal discussion concerning broad and general problems of interest to the board and the superintendent. Such a lack of dialogue could be related to conflict and estrangement that might be avoided by a presentation, in an informal setting, of the arguments and data behind some types of policy. In some cases of concern both superintendents and board members might be benefited by this dialogue. As an example, 73 percent of the board members and 61 percent of superintendents believe in the "teaching" of the desirability of the American way of life.[11] Such high approval upon such a question may indicate a lack of concern for the implications of such an approach. More frequent discussion of issues initiated by administrators would change policy positively. Also the discussion would possibly increase the board respect for the professional character of the administrator and his knowledge of the complexity of school operations and policies.

[9] Ibid.

[10] Neal C. Gross, Who Runs Our Schools (New York: Wiley, 1958), p. 116.

[11] Ibid.

The Teacher

Much of the impact of school administration and control is upon the classroom teacher. If the teacher perceives the administrator both as hostile and as the source of inadequate planning as one older study indicates, then the reaction of the teacher may be hostile.[12] Turnover rates or other objective measures of hostility plus attitude measures would be indicators of the adequacy of the administrator-teacher relationship. Since the nature and style of control may be the source of the hostility, means must be found to evaluate these relationships.

Related to this issue of morale and control is a parallel question of professionalization of the teacher. The administrator is a participant in a hierarchy that will necessarily limit professional autonomy but the teacher is not in such a role and may expect the degree of autonomy that the role model for the professional receives. The role model may be the college teacher. If this is correct then there is a crisis in legitimacy, for there is little indication that school administrators see the role of the teacher as being much above that of an ordinary employee. One writer on this topic said:

> The democratic concept of supervision emphasizes respect for personality. The teacher as an individual is to have freedom to think her own thoughts, to exercise initiative, to develop self-reliance and to assume responsibility for and intelligent participation in the direction of instruction and the determination of policies.[13]

While this statement may not be in itself objectionable it is in terms of obligations and not rights or prerogatives indicating authority. At another point the same author indicates a closer supervisory relationship:

> Although there has been much development of inservice programs for teacher education in recent years, there is still a great need for adequate supervision of the novitiate's efforts, and for close and careful suggestion, assistance, and direction so that he may develop desirable teaching methods and procedures, as well as broaden and increase his academic and professional knowledge.[14]

---

[12] A. S. Barr and Nels O. Reppes, "The Attitude of Teachers Toward Supervisors," Journal of Experimental Education, 4:237-301, June, 1936.

[13] Charles E. Boardman et al., Democratic Supervision in Secondary Schools (Houghton Mifflin, 1953), p. 28.

[14] Ibid., p. 8.

At another point the same author indicates that any delegation of authority is provisional. Other writers are even more blunt. One said, "It will be delegated by the group to a person or persons and withdrawn if not exercised toward the achievement of commonly determined objectives."[15]

If the content of a teacher's classroom performance were outside of the proper jurisdiction of administrative authority then the extensiveness of that authority would be blunted at an appropriate level. There is no indication of such a limitation. One writer of a general liberal persuasion argues that no teacher should be allowed to make derogatory remarks about the United States government.[16]

If there exists a conflict between professional and hierarchial perspectives on teacher roles or if there is a functional advantage in moving toward a professional perspective, then an evaluation instrument should assess the issues around this topic. Here are some examples of questions that are appropriate at this point: Is the out-of-class time of a teacher scheduled by school administrators? Are teachers required to participate in extracurricular programs? Is there a check made of lesson plans? Do administrators attend a teacher's class without permission? Are the class activities of a teacher defended by an administrator as being part of his authority? Are teaching methods determined by the individual teacher or teacher groups? Can a teacher invite outside speakers without permission? What are the tenure protections of teachers? What authority does a teacher have over the subject, age level, and time that he or she shall teach? Is attendance at meetings and conferences voluntary or compulsory? Are promotions based upon objective standards?

These kinds of questions are related to the degree of professionalism. There are other subjects that need evaluation in relationship to the teacher and controls. One of these is the degree of acceptance by the teacher of existing limits. Another is the participation in the making of policy. Yet another would be the degree of flexibility of policy itself. One last subject would be the informal compliance or noncompliance with policy.

The Student and Controls

The theoretically almost absolute and final power of the parent when transferred to the school system has left the impression that the issue of authority over the student is settled. Students lose most of their rights as citizens when they walk onto school grounds or participate in school activity. The issue is not settled and there are arguments for opening the issue. One argument is that such extensive control relations may not serve educational goals because they would encourage dependency. Also there would be a great discontinuity between high school and college where citizenship rights are beginning to be protected. Another reason is that the courts are gradually curbing some of this school authority. Also it is difficult to demonstrate democratic values in an autocratic situation. Another argument involves changing family patterns. If there is greater independence and self-reliance in the home there may be tension caused by a conflict between the home and school role expected of the student. Another factor involves intergenerational conflict. A teacher exercising unilateral authority may alienate and misunderstand another generation if the relationship to that generation is autocratic and thus somewhat disassociated. Such alienation would inevitably reduce communication effectiveness. For these reasons and others, the argument that is offered here is that control relationships with students, involving degrees of discretion, should be evaluated. Changes and modification resulting from evaluation might be advantageous.

---

[15] William H. Burton, "A New Definition of the Function of Supervision," Educational Methods, 18:6, October, 1938.

[16] E. Edmund Ruetter, The Administrator and Subversive Activities, (New York: Bureau of Publications, 1951), p. 105.

An evaluation could involve a range of issues. Following are some examples of issues: Are there clothing or hair regulations for students? Are these regulations written and distributed or are they assumed to be in the general authority of the faculty? Do students participate in writing the regulations? Are students consulted about regulatory policy? Are there appeals from disciplinary action? Are students involved in disciplinary enforcement? At which grade levels? Is the student government given any real authority? Is student government sort of a company union, or agency of administration, or is it a genuine lobby or policymaker? Are student organizations run by students? Do faculty sponsors have a veto over funds and activities? Is greater freedom or authority given older students or more advanced students in comparison to younger or less advanced students? Are students permitted or encouraged to disagree with teachers or administrators? Are reasons given when policy is announced?

These questions only indicate a few aspects of student control or regulation. Each of these deserve attention even if the decision is made to retain the status quo. This conclusion stands for the entire paper. The whole subject of school control should be a central one for school evaluation because it is central to the entire educational mission.

V

EDUCATIONAL EVALUATION: ADMINISTRATIVE FUNCTION

Troy A. Womack*

> Evaluation involves judging the worth of an experience, idea, or process. The judgment presupposes standards or criteria.[1]

Evaluation as a concept, as defined above, is as old as education itself. One can hardly conceive of formal educational processes occurring without this evaluation taking place, at least informally, in the minds of pupil, parent, teacher, administrator, and community observer. In fact, educational libraries are replete with volumes on "evaluation" of students, schools, curricula, buildings, organizations, teachers, school systems, administrators, and all other aspects of the educational milieu. Educators, students, teachers, parents, and communities continually "evaluate" the educational processes from their own frame of reference, and their behavior is influenced in the exercise of this procedure. Though most of this continual evaluation is informal, some of it has been formalized in the form of pupil reporting, teacher ratings, school self-studies, et cetera.

Evaluation in Business

Management theorists or administrative theorists have long utilized evaluation as an integral and basic link in the management or administrative process. McFarland comments on the administrative function as follows:

> Executive work is classified into three primary executive functions: planning, organizing, and controlling.[2]

He later portrays the interrelationships between these functions as illustrated below.

---

*Vice-President for Financial Affairs, Houston Baptist College.

[1] Paul L. Dressel, Evaluation in Higher Education (Boston: Houghton Mifflin, 1961), p. 6.

[2] Dalton E. McFarland, Management Principles and Practices (New York: MacMillian Company, 1958), p. 45.

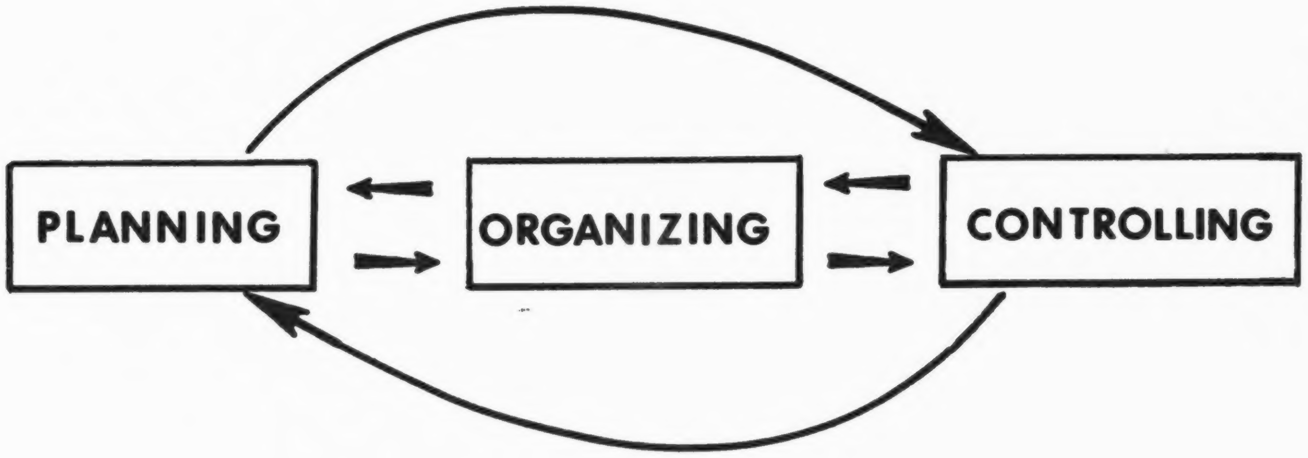

Figure 1

Reciprocal Relationships Between Planning, Organizing, and Controlling Functions[3]

Lehrer elaborates on the administrative process by listing four functions in the cycle:

1. Planning
2. Organization
3. Execution
4. Evaluative and Corrective Action[4]

Then he adds this comment, "This cycle is repeated in any management activity."[5] Lehrer's fourth function spells out the control function as twofold, evaluation and corrective action, though evaluation may be inherent in any control system.

Educational Organization

Dressel has adapted the administrative models of other organizations to educational organization. In a study of the stages of development in educational programs, he has proposed the following model:[6]

---

[3] *Ibid*., p. 299.

[4] Robert N. Lehrer, *The Management of Improvement*, (New York: Reinhold Publishing Corp., 1965), p. 7.

[5] *Ibid*.

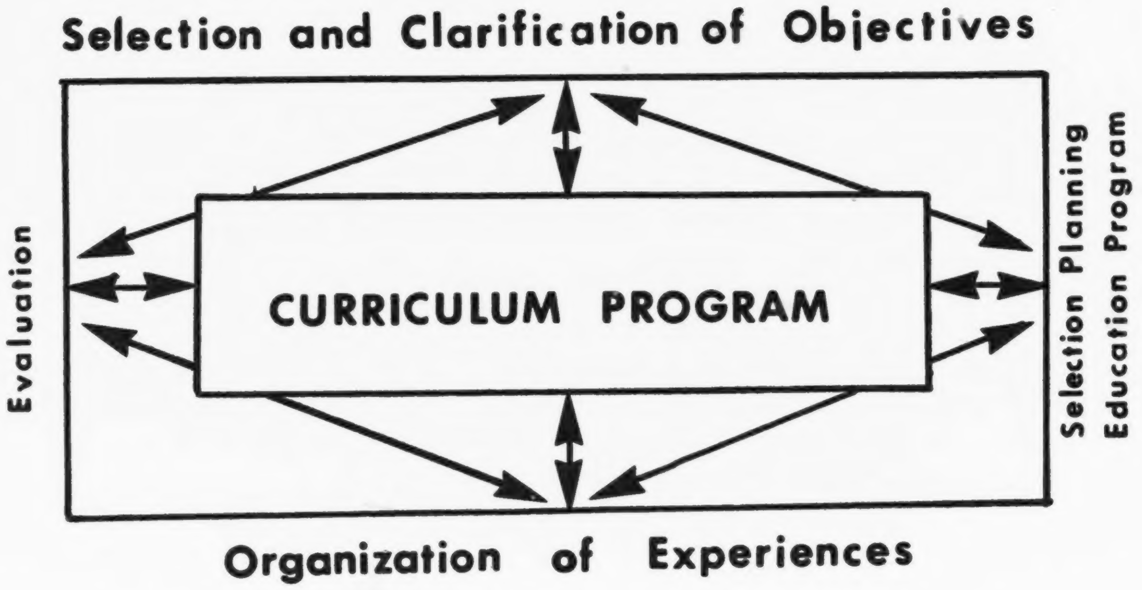
Selection and Clarification of Objectives
Evaluation
CURRICULUM PROGRAM
Selection Planning
Education Program
Organization of Experiences

Figure 2

Stages in the Development of an Educational Program

This illustration shows a close relationship to the ideas previously expressed by McFarland and Lehrer.

The relation of the administration of educational organizations to the administration of other organizations has been recognized as a mixture of similarities and differences.[7] Campbell, *et al.*, have raised four probing questions that might be asked of all organizations in order to reveal differences:

---

[6] Dressel, *op. cit.*, p. 9.

[7] Talcott Parsons, "Some Ingredients of a General Theory of Formal Organizations," ed. Andrew W. Halpin, *Administrative Theory in Education*, pp. 166-185. Also: Jesse B. Sears, *The Nature of the Administrative Process* (New York: McGraw-Hill Book Company, 1950).

1. What is the service which the organization is designed to provide?

2. What is the nature of the activity in which the organization will engage to perform this service?

3. What are the characteristics of the people who work in the organization?

4. How may the activities of the organization be appraised?[8]

In reference to their first question, two continua were suggested.

Cruciality to Society . . . .

Least Crucial → Most Crucial

Factory Fire Department School

Public Visibility and Sensitivity . . . .

Least Visibility and Sensitivity		Most Visibility and Sensitivity
Factory	College	School[9]

Two continua were suggested relating to the second question.

Complexity of Function . . . .

Least Complex → Most Complex

Factory: making bolts	Government: collecting taxes	Schools: teaching children	Psychiatric Staff: changing personality

---

[8] Roald F. Campbell, John E. Corbally, Jr., and John A. Ramseyer, *Introduction to Educational Administration*, (Boston: Allyn and Bacon, Inc., 1962).

[9] *Ibid.*

Intimacy of Necessary Relationships . . . .

Least Intimate ⟶ Most Intimate

Factory: making bolts	Government: collecting taxes	Schools: teaching children	Hospitals: treating the ill[10]

Regarding question three, the following continuum was presented:

Staff Professionalization . . . .

Least Professionalization ⟶ Most Professionalization

Factory	Hospital	School	College[11]

And regarding the last question, this continuum was given:

Difficulty of Appraisal . . . .

Least Difficult ⟶ Most Difficult

Sales Organization: sales	Manufacturing: Organization: products	School: change in behavior	Church: inner and outer change[12]

It is suggested here that "appraisal" as used above is synonymous with "evaluation" as previously defined.

These interesting continua point up some characteristics of educational organizations that are different, but they do not appear to invalidate the generalization that human organizations have a great deal in common. If the different human organizations are similar in operation and development, then it may follow that administrative strategies and theories utilized by other organizations may be adapted for use within the educational organization, and that the concept of evaluation which is inherent as a part of the concept of control in administrative theory may also be a keystone or integral part of the administration of educational processes, on the part of teachers, administrators, and institutions.

---

[10] _Ibid._

[11] _Ibid._

[12] _Ibid._

At this point, the connotations of such terms as administrator, controller, evaluator, or leader need not be a concern to educators, for the educators--teachers, principals, superintendents, professors, counselors, or others--are all of these in their own classroom, building system, course, or group. The most advanced management theorists--Mayo, Likert, Hertsberg, Argyris, and McGregor--recognize the power and the efficiency of the democratic group dynamics in the achievement of productivity, particularly where educated and professional people are involved.[13]

Writing on the strategy of modern managers and administrators, Gore concluded:

> At the core of this strategy is the relation between two of the logician's favorite concepts, means and ends. First, broad purposes are broken into fragments defined to serve as components of the desired ends. These ends in turn are translated into means appropriate to fulfilling them. . . .The strategy has a second phase whose means, defined as a proposed course of action in terms of the objectives sought. . . are translated into accumulated individuals' patterns of activity that will fulfill the objectives defined as a means. When this is given form and substance by the insertion of people, materials, and facilities, a rational system of action is established through which the ends initially formulated as collectively sanctioned values are realized.[14]

Gore, as a political scientist, comes to the same broad conclusions as administrative theorists, management practitioners, and at least some educational administration theorists: i.e., that the achievement of goals through people involves the setting of objectives--the ends--and the planning, organizing, executing, and controlling based on feedback and evaluation--the means.

Evaluation Functioning in Education

The function of evaluation in education then may be to use the familiar and comfortable concepts of evaluation as now practiced in education as a window or entry-way into a process or strategy that recognizes and utilizes objectives--planning--organizing--evaluating--controlling cycle as a useful and necessary one for progress and self-control of education.

---

[13] Saul W. Gellerman, Motivation and Productivity (New York: Vail-Ballou Press, Inc., 1963).

[14] W. J. Gore, Administrative Decision-Making (New York: John Wiley and Sons, Inc., 1964), pp. 3-4.

The establishment of criteria, goals, or objectives is no innovation in American education. "The Quality Measurement Project in New York State"[15] and the current "Pennsylvania Plan"[16] are examples of state-wide efforts in this area. Countless examples in individual schools and classrooms could be cited. Drummond, writing in the Association for Supervision and Curriculum Development Yearbook of 1960, states,

> Criteria stated or implied are essential in the process of appraisal or evaluation. The extent to which values deemed desirable are supported provides both a description of present status and identification of areas in which future efforts are needed.[17]

Perhaps the goal emphases have been limited to measureable knowledge or intellectual achievement, overlooking some vital social and psychological measures.

Drummond states the difficulties of evaluation in education succintly when he writes,

> Use of appraisal techniques to improve leadership for instruction is difficult because educational goals are multiples and teacher-learner processes are complex. . . The complexity of the teaching-learning processes can easily be demonstrated.[18]

The problems of application of management or administrative theory including the process of evaluation to education are numerous and complex, for example: (1) agreement on operational educational objectives, (2) complexity of the educational process, (3) difficulty of defining and classifying the teaching-learning processes, (4) development of demonstrable educational models for different processes and situations, and (5) difficulty of valid and reliable measurement of objectives, factors, and results of the educational process.

---

[15] William D. Firman, "The Quality Measurement Project in New York State, Science Education, 50:259-279, April, 1966.

[16] Henry S. Dyer, "The Pennsylvania Plan," Science Education, 50:242-248, April, 1966.

[17] Harold D. Drummond, "Appraisal: A Method for Improving Leadership," Leadership for Improving Instruction (Washington, D. C. : Association for Supervision and Curriculum Development, 1960), p. 162.

[18] Ibid., p. 163.

In spite of the difficulties involved in the magnitude of the problems, our innovative and progressive society living in an unprecedented era of knowledge explosion coupled with affluence, may demand that an attempt be made to apply this previously discussed cycle of administrative theory to education in the light of the cruciality of education in such a world and the possibility of improvement by such an application.

VI

PROGRAM EVALUATION IN EDUCATION

Wallace H. Strevell*

The concept of program evaluation is a logical phase of program budgeting. The assumption in program budgeting is that all the work of an enterprise can be described in several programs. These programs are accomplished by various and variable activities, the achievements and effects of which may be observed. Evaluation of an enterprise would thus be based on evaluation of its programs; and evaluation of any program would be the sum of the achievements and effects of its several activities at any time. Cost-utility analysis applies particularly, therefore, to the costs and achievements of each activity within a program.

This evaluative approach seems on the surface to be systematic, direct, and adequate, which for many purposes it may well be. In order that a clear interpretation may be given, however, the concept of program budgeting itself needs to be explored. A stimulating description of program budgeting was prepared by Rand Corporation in 1964.[1] The following year the planning-programming-budgeting system (PPBS) was introduced at executive request in several departments of Federal Government. The conceptual framework of program budgeting is adaptable to education, federal health expenditures, natural resources activities, transportation, space programs, defense, and other enterprises of Federal Government. Since the system appears to serve useful purposes in any governmental department, it probably will after awhile be accepted for the management of state and municipal budgets as well.

The language has crept into school system planning in such forms as "a Title I program", or "a program for disadvantaged youth." However, the school practitioners may not yet realize that the word program has acquired a technical meaning. It requires an evaluation system that will feed back into future budget planning in terms of relative cost effectiveness, or in other words cost-benefit management. In its planning feedback phase, program budgeting is of interest to many parties: the local community, the school authorities, the approving agencies, and the recipient of services.

---

*Professor Educational Administration, University of Houston.

[1] David Novick, (ed.), Program Budgeting, Superintendent of Documents (Washington: Government Printing Office, 1964).

Nature of Program Budgeting

The conceptual framework of program budgeting is mission oriented, encompassing all activities and responsibilities of the enterprise. As stated in the introduction to program budgeting for the City of Baltimore,[2] the primary objective "is to make available to all interested persons, in clear and concise form, the programs of work and services to be rendered, together with their costs." Thus a program in program budgeting is defined as the work to be accomplished. Since the program level generally will be the appropriation level, careful forethought and judgment should precede adoption of program titles for an enterprise. In education the programs adopted would represent broad areas of services to be rendered or produced for the public good. All the work of an educational enterprise logically would be performed within the scope of its several program titles. Traditionally public school budgets have not been organized in this pattern. Yet the traditional school budget format is not incompatible with a planning-programming-budget system since costs of any activity can be distributed and accounted under as many character or object classifications as desired for purposes of reporting. Program budgeting demands that the educational work and services be described more clearly in terms of utility and benefits than the traditional school budget format provided.

Having defined the work and services in terms of all the programs that make up an enterprise, the next step in program budgeting is to state the activities planned under a particular program and the anticipated costs of each activity. Budget planning for subdivisions of programs in essence specifies how a program is performed. The activities stated may be expected to change from year to year or even month to month in some respects. Some activities will cost more than others; some will be of longer duration than others; priority of activities may be affected by needs, resources, or available skills; many activities must proceed according to schedules that will be affected by another activity. Considerable cost-utility judgment and flexibility exist in the decisions as to planned activities within a program. For example, The Gulf Schools Supplementary Education Center has in its charter five programs: (1) Administration, (2) Planning Services, (3) Evaluation Services, (4) Dissemination, and (5) Application Services. The activities listed in the Evaluation Services Program are: (a) Publish position papers of national relevance, (b) Develop models for evaluation, (c) Train a leadership corps in evaluation, (d) Demonstrate process evaluation, (e) Design evaluative forms for collecting data, (f) Write evaluation reports for clients, (g) Conduct a continuous evaluation of GUSSEC, (h) Produce evaluative measurements instruments, (i) Provide simple statistical treatment of data for clients, (j) Produce descriptive statements of projects, (k) Provide a clearinghouse of information on evaluation methods.

---

[2] City of Baltimore Budget Manual, Department of Finance (Baltimore: Bureau of Budget and Management Research, 1966).

The budget allocation for Evaluation Services is prorated to these activities, as functions of expenditure. The planned costs for each activity are distributed by traditional object categories. In summary, the budget format is directly related to planned accomplishments.

Challenge of Program Analysis

The present practice of attaching programs to already existing enterprises encounters several planning and evaluation issues. Thus we have the local public school system as an existing major enterprise of long standing and constitutional tradition, to which there is being attached a cost-benefit management type of program funded in part by the Federal Government.[3] Let us say the school system does and for a long time has instructed children in the skills and choice of reading, but now it purposes to attach a federally funded Title I program in remedial and enriched reading for disadvantaged children. While this is quite possible, probably involving a cost sharing of local and federal (as well as state) resources, the issue of evaluating the accomplishments of the federally funded program in meaningful federal program budget terms becomes complicated. The natural impulse in this case is to measure the gain in reading achievement of the disadvantaged child, but to do so is actually to measure the product of the whole school enterprise. In order to have benefit data with which to evaluate and plan attached federal education programs for the future, may it not become obligatory to appraise the whole existing educational enterprise and thus to direct and control the whole enterprise? Note that the cost-benefit management program is not an independent program but is attached to and depends upon an existing enterprise. The conclusions may well be either that the existing school enterprise should adopt PPBS, or the attached programs ought not to be evaluated, or that ways and means should be sought to evaluate the attached federal programs solely in terms of their own activities without including the effects of the enterprise to which they are attached. If the latter course were adopted, it would be meaningless to measure the gain in reading achievement of affected children, but the only justifiable course would be to measure the specific experiences that the children had as a direct result of activities conducted by the Title I program. The benefit for cost-analysis purposes would be only what the Title I program accomplished.

Public school practitioners are not accustomed to accountability for their educational programs. The evaluative reports normally prepared for professional use and for public relations are obscure and tenuous. This is explained by inability to measure the whole product, or by the complex influences of community and school on learning. Schools render good accountability of strictly object-related costs such as time reports of employees and auditing of invoices. Cost-utility analysis has occurred in auxiliary services such as transportation, cafeteria, and plant maintenance. But results or effects of the educational programs are more often inferred from such factors as teacher competence, modern methods, staffing ratio, and curriculum content, which generally are claimed to represent school quality or excellence. Attempted measurements of academic and other learning achievement have proven not only to be inadequate but more important to be relatively unrelated to the actual program. Today, if applied to school systems, the planning-programming-budget system offers a better promise of adequate accountability and management.

---

[3] Elaine Exton, "Word from Washington," series in American School Board Journal, November and December 1966; Janaury 1967.

This application of evaluative methods to planned activities in school budgetary programs is the challenge of program evaluation. Obviously, planned activities are multiple; they are scheduled to start, proceed, and complete at various, often overlapping times. Their priority of support and emphasis is variable. Even the nature of an activity itself will change during the process. Futhermore, the component activities of a budgetary program represent a variety of goals and of ways and means. Some degree of conflicting purposes and objectives among the acitivities can be tolerated, such as competition between activities, or different views held by the personnel conducting different activities. Planned activities are subject from time to time to outside forces. New objectives can emerge in the process of an activity. The achievements of an activity are related to available resources, skills of personnel involved, and quality of leadership. The scope of an activity may be changed or limited by a changing concept of the institution or enterprise as a whole.

Systematic Approaches to Evaluation

In general, the approach in evaluating an activity is to determine the kinds of information demanded, the method of discovering relationships among the observed data, and the basis for interpretation of the information. A commonly applied system is Objectives→Results. This method assumes that the stated objectives are complete and openly expressed, that no concomitant objectives will develop during the process, that all the value-system needed for appraisal is contained in the statement of objectives, and that such an evaluation would be useful in future planning. It seems to be implied that the end justifies the means, and that effects can be disregarded when the results are known. Measuring results in terms of objectives will probably continue as a popular system of evaluation because of its seeming simplicity, but obviously it opens the door for all kinds of uncontrolled side issues.

A more recent system of evaluation is Input→Operations→Output. This approach is advantageous as a research design, since one may vary either Input or Operations and observe the change in Output. The system is favored in operations research because it enables feedback during the process. When Output does not meet changing needs, for example, the Operations can be adjusted or the Input varied. An illustration in educational analysis would be to increase resources (the Input) until a significant improvement is observed in learning (the Output). This appeals to those interested in scientific approaches since engineering scales of predictability can be established.

Another different system is →Measurable Effects←. The implication here is that acceptable criteria are available by which to compare and determine the desirability of the outcome. Assuming that a value-system is adopted and relevant observations of the outcome are completed, this particular system of evaluation does not exclude any of the countless forces, influences, internal stresses, conditions, or activities in the total field that might relate in some way to the observed effect. The Effects system is highly sophisticated, but the task of reducing to mathematical terms the massive data that describe all field forces assumed to produce the effect is almost prohibitive. The system invites much speculation and formulation of hypotheses as to the unknown forces. Most of the speculation escapes proof, since the observed effects are mainly situational and could be only approximately replicated for scientific study.

An evaluative system that lends itself especially well to management decision-making, particularly in planning and budget-allocation, is Mission →Achievement (Additive). The emphasis here is on hard evidence that justifies adoption of practical operating plans. Examples of such evidence are: (a) Conclusions as to more efficient and economical ways to achieve desired results, (b) Critical-Path-technique evidence as to effective scheduling or priority of activities, (c) Predictive data as to public demand, needs, or welfare.

Timely Implications

In the operations analysis program of the U. S. Office of Education, the current research on techniques of evaluation has been directed toward creating representative models, generally based on Input-Output principles.[4] Some models are mathematical or symbolic such as cost-benefit research; other models are simulative showing associated factors and processes, for example, all major factors related to student achievement. Inevitably these models are not confined to what federal projects themselves produce, but as explained above, they encompass the whole impact of federal projects on the American educational system. While these research models are no more than abstract statements seeking to describe relationships, sets of variables, and parameters, the standardizing effect of such positive evaluation of decision-making in funding appropriations can be sufficient that local school districts could well direct attention to other justifiable principles such as diversity.

---

[4] *Ibid.*, January 1967.

Considerable professional energy will need to go into operations analysis at the local level before school systems can produce evaluative models indigenous to their own local enterprises. The school practitioner is so unaccustomed to program accountability or cost-utility evaluation, that school districts generally do not have the professional staff trained or assigned to develop local evaluative research. Larger school systems could take the initiative in this regard by updating the training of their budget analysts, and probably by experimenting with program budgeting, or even innovating a mission-oriented building unit type of budget. Attention of chief administrators in local school systems should be directed toward the development of models utilized in evaluation that would reflect the values and needs of a well-established, responsible, ongoing enterprise.

VII

QUANTIFICATION OF EDUCATION FOR PLANNING APPLICATIONS UTILIZING MODELS FOR EVALUATION

Glen Self*

The role of evaluation models for the educational process by nature must be elusive and controversial. The reasons that this is true are the same for any process that is modeled in the management area or in those areas where physical measurable data are not available. Therefore, the controversy surrounding evaluative models for the educational systems is not a new perspective in modeling that has not been considered previously. The difference between this type of model and a model that describes the drag on a lifting body is that there is some emotion involved. The quantification of that emotion becomes a necessary part of the considerations in the construction of a model of this type. Before becoming involved in the techniques, consider the role of the utilization of models for the evaluation of educational functions and the philosophy behind their development.

Model - Evaluation Relationship

The definition of a model will be any logical flow of information that is designed to provide the decision-maker with a better perspective of the data and that information necessary for making decisions. To be more definite, a model will be restricted to only those functions which can be converted to a computer code and subsequently processed on a computer. This is equivalent to stating that it is necessary to be able to explain all aspects of a model to a level of detail sufficient for any person to carry out the functions of the model. This does not in fact eliminate the person or the decision-maker from having a part in the function of the model. It should be pointed out that the information supplied a model is always in quantitative form and the ultimate use of the output of a model does not include the decision that is made and in effect does not force the model user to quantify his decision in a numerical sense prior to making the decision. Inputs are not of this nature. If inputs are deliberately withheld from the model due to their subjective nature or the emotional nature of their effects, then they must necessarily be part of the decision process. Often the construction of a model without the incorporation of this type of information is quickly discredited by a specific counterexample. One mathematical argument that model users everywhere understand is that a single counterexample can be used to disprove a model. In an area such as evaluation of educational process, products, systems, etc., it is going to be necessary to avoid pitfalls of this type if a generalized evaluation model is to be formulated.

---

*Assistant Professor of Industrial Engineering, Texas A & M University

A generalized evaluation model will necessarily provide for a device that can be used as a planning tool and as a training device for those persons concerned with the problem of education and the evaluation of education. It will be toward this that the remainder of this paper will be addressed.

Model Building Approach

The model as discussed here can become an important frame of reference among persons discussing evaluation and the effects of plans upon the evaluation of a specific program. This is possible due to a model being intransitive under discussion as opposed to opinion or written works which are subject to interpretation in the absence of their author. It is assumed that with proper background in basic logic and mathematics a planning model can be understood without qualification by any person. The relative importance of various aspects of the model are defined exactly and the model builder need not be present for the model user to understand exactly the premise for occurrences and relationships among variables within a model. Further, a model will permit analyses to be duplicated by individuals remote to each other. Admittedly, interpretation of the results may differ; however, the results of the model are reproducible.

In order for evaluation procedures to gain widespread acceptance it will be necessary for analyses to be reproducible by following documented methodologies that yield exactly the same output with identical inputs regardless of the locale or the person using the model. Therefore, the resolution of semantic interpretation and emotional differences is an important aspect of reducing evaluation procedures to a model format. Even if differences occur over the model itself, discussions can be based on the quantitative terms of the model.

The construction of a model for a field such as education evaluation introduces another aspect that has come to all other fields which have widespread accepted use of models; that of determining what data should be collected and the proper format for collection of those data. The early stages of construction are generally concerned with the reconstruction of historical data in order to provide a baseline from which relative comparisons can be made. This has often proved to be the only worthwhile results of three or more years of concerted effort in equally complex fields. However, it is part of the down payment on progress toward a useful model, and in its absence a good model will never be constructed, at least not within

2

any near time period or for use in projection type applications.

The third area of model building which is becoming of increased importance that of quantifying the experience backgrounds or expertise of experts in a field in order to supplement the measurement type data that is available for model building. Recent advances or trials in this type of approach to model building may be found in the bibliography. The basic philosophy of this type of an approach is that all elements of the model must be established in terms of numbers. That is, all elements must be placed upon a cardinal scale if they are to be included as part of the model or the decision-making process. This is consistent with decision theory techniques developed in recent years.

Dynamic Models

There are at least two major divisons for planning or evaluation models, static and dynamic. As the names imply, the basic difference is whether the model responds to time-dependent information of whether the model provides the same output regardless of the time-dependent processes operating. Often the dynamic model is one that is unaltered during its execution phase in that the decision functions, the initializing values, and the time-dependent functions which describe process responses are imbedded within the model structure.

Generally the planning model should be of the type that reflects the actual planning situation, especially in those cases where part of the input information is subjective in nature and subject to interpretation. This is not the case in the normal static type of mathematical model which is used to model physical systems. The role of the model in those cases is greatly different from the one that will be used in evaluation of educational programs due to the availability of measured data. The physical process is repeatable and the prediction of outcomes based upon those physical processes do not change from locale to locale or from experimenter to experimenter. Therefore, the use of a mathematical model in the planning environment or in the evaluation of educational programs should be the type of model that represents the physical conditions in time and does not permit the look ahead type of optimization of a system that will make decisions and judgements without specific directions from the model user.

The evaluation model should be limited to planning conditions such that the states of the various variables are initialized and projections of model are made for some finite amount of time into the future. The decision functions relating to those variables are specified in advance of exercising the model. The actual mechanics of the model provide for it to move from one time period to the next time period, making all decisions and displaying the results obtained from initial input and model generated information. This type of model must be constructed such that it is regenerative in nature in that the initialization data are simply a starting point and when those data are exhausted the output of the model will continue to be generated by the mechanism defined as part of the model.

The worth of the planning type of model is that the output from the model must represent the real-world process in a continuous fashion, thereby providing a basis for judging the internal structure of the model as a fair representation of the actual process. Through this type of modeling technique the evaluation of an educational system can be performed by simulation. If the model can be constructed such that it represents the present system in the absence of the influence of the initial input data, then there is hope that the parts of the system and the relation of those changes to the other parts of the evaluation process may be determined. In an area highly dependent upon subjective data, this seems to be the only reasonable approach to evaluative model for school systems.

A premise which is basic to the construction of a model of this type is that the process can be described in quantitative terms. In support of this viewpoint it is suggested that a model is not conducive to communication or general use unless the results are repeatable through the methodologies associated with the model. If identical results are not obtainable from identical data that are supplied the analyst, then the model has only a single application.

The conclusion to this phase of the discussion is that the structure of the model should be such that the person evaluating an educational process could utilize the basic model structure for analysis of his data and be able to determine the worth of his objectives for his program relative to his own state or initial conditions without being thrown into a rigid evaluative scheme that is necessarily biased in favor of his own particular state or against it. Therefore, the continuous or dynamic model that does not permit look-ahead type of capability is recommended due to the fact that it places the model user in a more nearly real-world type of situation than does the conventional regression type of models that are by their nature static and highly subject to interpretation external to the model. For the type of model discussed above, the interpretation is part of the model and is output in an interpreted form in order that discussion based on a common frame of references can be initiated among the discussants. Further, the type of model described is considered to be more equitable than most other types of models suggested in the literature.

Utilization of Subjective Data

The need for expertise-oriented models is demonstrated by the rejection of recent statistically oriented studies. The tendency is to qualify the statistical analysis with conditioning statements which make the results palatable. The point is that these qualifications have not been quantified and placed in a proper context to the statistical studies.

Based upon the present literature for evaluation in education, the evaluative process is generally concerned with one of the three areas described as the structure, process, or product. It seems reasonable from the model-building standpoint to be able to consider all three aspects simultaneously or at least by interfaced models. The normal tendency appears to be one of making application of statistical type analysis on the numerical data associated with one of these three areas. These analyses are then interpreted in terms of an individual experience base. It is contended that the analysis of the data has been held back by the failure to introduce the subjective type of data in a quantitative context and to use that information as a bond among the three areas.

In terms of the normal model-building process, the present area described as structure of the educational system will provide the input data in terms of student to teacher ratios, number of classrooms, teacher salaries, age of facilities, worth of facilities, total enrollment, etc. These are all state variables which can quantitatively describe the present environment of the educational process. The actual internal mechanics of a model can be determined by what is presently called the process type of evaluation. This is possible since the process represents the policy type of decisions that are used to determine the type of counseling, testing, and measurement that the student is to receive in the educational process. Therefore, it is felt that a generalized model can be constructed that will be representative of most of the types of polcies or actions that may be observed in the processes that the student passes through in his educational experience. The output of an evaluative type model should describe the third area, that of the product. The product is measured by the achievements of the students that are produced by a specific educational process. These studies generally take the form of longitudinal studies which evaluate the product which infers that measures of education are available in forms other than achievement test scores. Thus, at this point in time it would appear that an oversimplification of the educational evaluation problem is one of bringing the structure, process, and product type evaluations together in a single comprehensive model.

If it is generally agreed that some relationship must exist, and that it is desirable to model that relationship, then the data that represent the interrelations must be established. It is possible that the data do not presently exist to accomplish this. If at all possible, experiments should be originated to collect the applicable data. It is anticipated that some of the data will be of type discussed above, quantification of expertise.

An example of the type of quantification that is being considered is as follows. Assume that it was decided that the number of grade points received in civics courses in the student's secondary educational experience is related to the post educational activity in civic affairs in the community and the state. Performance in those activities is considered as a measure of the product of the educational process. The measure of this is hardly a directly assessable quantitative measure of the product. However, through a number of statistical computations it is probable that a cardinal index could be determined for the purposes of building a model. A utility measure of the value of civics relative to the student's performance could be determined directly through the use of expertise or at least established by the model user. By this method the evaluation used by the model could be displayed for all persons who review the evaluation procedure . This would provide a point of discussion and comparison among different experts and model users, but all would be referring to the same quantitative value. This is a relatively simple example which can be expanded. For example, consider the question of the interaction of the college level activities and education with the secondary education relative to the role of civics and its effect upon the product (student). The adjustment for this interaction could be carried out by the formulation of an amplification function which describes the postgraduate contributions as a function of both the secondary education and the college level education. By utilizing functions of this type, derived by mutual agreement of the model users, two schools could be compared on the basis of the civic activities of the postgraduates regardless of the structure or additional education received by the individual. This type of modeling would tend to reduce the bias by allowing the model builder and model user to consider each aspect of the evaluation prior to using the model to combine those considerations in the form of an overall evaluation of the process.

The necessity of introducing subjective data is assumed to be already established, otherwise relatively complex models would have been developed as a result of the amount of attention that has been placed in evaluation of education. A more basic reason for systematically introducing the subjective data is that the results of the model must be reproducible from standard instructions. In addition, a major complaint about decision-making models, evaluative models, and planning type models is that the model generally represents the point of view of the model builder. By using the approach discussed in this paper, the model will represent the point of view of the model user. The model user is responsible for the results obtained from the model; therefore, he should have the major share of the control as well as the responsibility for the results produced by a general model.

An illustration of this relationship between the model builder, the model user, and the model can be found in the classical, multiple, nonlinear regression analysis model from applied statistics. In this type of analysis the model is programmed for the computer and instructions are issued as to the inputs required. The model produces outputs in accordance with standard statistical procedures. The model builder is only responsible for the correctness of the model which does not include the input data or the final interpretation of the analysis. The input will necessarily include the functional forms for the variables in the analysis. Therefore, the responsibility of the output is that of the user. If the model builder had restricted the functional forms or the number of data points that could be used, then he would have a larger responsibility for the final analysis. Perhaps evaluative models are too restrictive and do not permit the expression of the model user to be reflected to a sufficient degree. It is proposed that more general models be developed using a systems approach. This is a general objective that is worthy; however, it should be realized that techniques will need to be developed before this can be accomplished. During the early formulation of an approach of this type, the point will not be as to whether the model is exactly correct, but will be that the results are reproducible and that the total flow of information can be referred to directly by discussants remote to each other.

Methodology

The methodology for the development of an evaluation model will follow the normal form of model development in that the desired outputs should be specified first. These outputs should represent the set of measures that are pertinent to the evaluation; that is, the interpretation should be completed such that the numerical values that are output appear on a cardinal scale as an evaluation of the structure, process, and product. The measures used will generally describe the product of the educational system. The output may be a single composite index of the evaluative process, a number of ratios which compare the various parts of the educational process, or some combination in between. The most important point is that the outputs defined for the model must represent the interpreted form of the evaluation if the model is to be widely used. Initially, the desired outputs should be specified without consideration as to the data that are available. This will necessarily be refined by the available input data and the formulation of the model.

The input data for an evaluative type model will be of two types; those which are needed to describe each individual problem that is proposed to the model, and those which are generally applicable across a range of problems and are not changed frequently during the running of the model. The second type of data are usually referred to as library data for the model. These data may be changed but generally less frequently. For example, the description of the student body, the staff, and the facilities will change with each new problem; however, national economic projections, national population trends, and some facility projections would be expected to be relatively stable for a number of runs of an evaluative model. Therefore, in order to reduce the inputs to the model each time it is run on the computer, it would be an advantage to have a single data bank which represents the stable factors. The development of these types of data by a study group would increase the consistency of the results obtained from an evaluative model. For widespread acceptance of a model the number of inputs must be kept within reason. There is a tendency in the systems approach to provide models which read in large amounts of data only to print them out without their entering into the analysis that was performed. Therefore, the level of detail in the inputs should be consistent with the level of analysis provided by the model and the level of detail required by the outputs. Caution should be taken not to require the model user to perform laborious computation prior to inputting the data to the model; that is, quantity is preferable in inputs over complexity.

There are two types of data that are used in the formulation of the model, measurable and subjective. The role of the subjective data in the evaluative model has been discussed at length in the above and will be summarized below. First consider the role of what has been defined as measurable data. Measurable data is that which already has its own cardinal scale that is acceptable to people in general and more specifically the potential model user. For example, Graduate Record Examination scores have an established scale, and while experts do not necessarily agree as to the meaning of the score, it is considered measurable. The basic need for data in the modeling process is to define relationships between inputs to the model and the outputs which it produces. Often, there are a number of relatively complex relationships that must be evaluated prior to the display of the results. The role of the data required during the formulation of the model is to provide sufficient experience in order that the proper relationship between the input data and the ultimate output values may be established. It is considered the obligation of the model builder to analyze the applicable measurable data prior to utilizing subjective data. Due to the wealth of data that exist in the measurable form that describes the education process, much of evaluative model can be constructed from existing data. The types of analyses used for raw data in determining relationships are an important part of the model building process. It is not true that the same data analyzed by different methods will yield the same results. Therefore, the method of analysis should be selected to be consistent with the process the data represent. For example, there is some discussion between the merit of regression analysis versus factor analysis for developing relationships among variables. A still more refined argument is between the use of step-up regression and step-down regression in model development. In this case there would be little difference if the analyst could order the importance of the variables on some basis independent of the analysis procedure. In general, factor analysis is a more liberal approach to the consideration of various combinations of variables than will appear in the relationships developed within the model.

The major class of techniques used in development of models using large quantities of data are curve fitting techniques. These are the fitting or hypersurfaces in the multidimensional cases and further increase the importance of selecting the appropriate methodology. The objective of the analysis of the measurable data is to reduce it to a usable form within the model. The data analysis may use other models to obtain the refined form of functional relationships desired, but once they are obtained, they will be used over the life of that particular model.

It is anticipated that the amount of measurable data as compared to the amount of subjective data that would be used in an evaluative model is relatively large. It is suggested that the normal data analysis procedures be used to establish all relationships feasible and that the subjective data approach be used to fill in those relationships which are required to provide the outputs. It should be realized that the subjective data approach required statistical analysis usually in excess of that required for the measurable data. The use of subjective data introduces new problems such as obtaining convergence of the estimates of the experts without introducing bias, selection of scales for measuring the subjectively determined data, conversion of analog data to digital form, problems of discrimination among questions, and distinguishability of objectives. There are problems that must be dealt with prior to the standard analyses that are applied to measurable data. Therefore, if the subjective approach is treated in a scientific manner, it is not an easy shortcut to formulation of an evaluative model. The contributions of subjectively determined data are usually the first to be scrutinized by the model user.

During the developmental stages of model formulation, it may be necessary to substitute the subjective data approach for data that are measurable in nature but not available. A program then can be instituted to collect the needed measurable data and replace the subjectively determined relationships when possible. By proceeding with the model formulation, using subjective data, more exacting requirements for the type and format of the measurable data may be established.

There is a limit on the level of detail that can be considered by a broad model of the type discussed here. This limit is imposed by the size of the computer available for running the model. Therefore, it is not proposed that the level of detail be at the individual student level, but that general classifications be developed for the various parts of the structure, process, and product that would make the model computationally feasible. In some cases the stratification of data in this manner will reduce the analysis required and expedite the formulation and programming of the evaluation model.

Conclusion

Quantification of subjective data has been associated with the field of operations research for a number of years in the area of utility theory. The experimentation in the determination of utiles or value associated with various types of outcomes has met with varied results. Perhaps the failure or success of these investigations has been dependent upon the quantitative tools or methods which have accompanied the subjectiveness. The rigor of statistics, mathematics, various optimization techniques, and significance tests have something to contribute to this area. The technique itself is seen as a logical development in the extension of complex models to approximate more closely the total information available to an evaluator. The intent of this approach is not to replace the decision-makers. It is the systematic consideration of all aspects of the problem, one factor at a time, and the structuring of the analysis such that each of the elements will be combined in the appropriate manner to provide the integrated analysis desired. Perhaps before judgment is passed upon this approach, the strictly measurable data approach should be tried for the model described herein.

The model building approach discussed in this paper has been directed toward the conversion of predetermined variables and desired outputs into a form that may be used on a second generation computer. The evaluation model is not to be confused with the information storage and retrieval type of computer program that is used in conjunction with a central data bank which keeps administrative records for a school system. The evaluation of a program will not change rapidly enough to require real time processing of information. It is entirely possible that the updating of the evaluation model by newly developed relationships will become a relatively frequent task, in which case the evaluation model should be the responsibility of a centrally located data center. Generally the evaluation of an educational program is the result of studied deliberation; therefore, an evaluation model, centrally located, could serve a large number of schools. Remote terminals and direct communication with the evaluation model is not foreseen as a real need in the near future.

Generally any model that can be used in a planning context can also be used as a training aid to demonstrate the outcome of various occurrences in the input data or in the model structure itself. The type of evaluation model discussed above should contribute to both aspects of the education evaluation task.

BIBLIOGRAPHY

A. Books and Pamphlets

Adams, Georgia S. *Measurement and Evaluation in Education, Psychology, and Guidance*. New York: Holt, Rinehart and Winston, 1964.

Bloom, Benjamin S. (ed.). *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain*. New York: David McKay Company, Inc., 1956.

Boardman, Charles E., *et al*. *Democratic Supervision in Secondary Schools*. New York: Houghton Mifflin Company, 1953.

Brumbaugh, A. J. "Evaluation of Schools," *Educators Encyclopedia*, Englewood Cliffs, New Jersey: Prentice-Hall, 1961, pp. 424-429.

Bureau of Budget and Management Research. *Budget Manual*. Department of Finance, City of Baltimore, Maryland, January, 1966.

Callahan, Raymond E. *Education and the Cult of Efficiency*. Chicago: The University of Chicago Press, 1962.

Chamberlin, Leo M. *The Teacher and School Organization*. New York: Prentice-Hall, 1958.

Cooperation in General Education: A Final Report of the Executive Committee of the Cooperative Study in General Education. Washington: American Council on Education, 1947.

Dressel, Paul L. (ed.). *Evaluation in General Education*. Dubuque, Iowa: William C. Brown Company, 1954.

_______. "Measurement and Evaluation," *Thirteenth Yearbook (1960) of American Association of Colleges for Teacher Education*. Washington: The Association, 1960, pp. 45-52.

_______, (and Associates). *Evaluation In Higher Education*. Boston: Houghton Mifflin Company, 1961.

_______, and Lewis B. Mayhew. *Critical Analysis and Judgment in the Humanities*. Dubuque, Iowa: William C. Brown Company, 1956.

_______, and Lewis B. Mayhew. *General Education: Explorations in Evaluation*. Washington: American Council on Education, 1954.

Educational Testing Service. *A Plan for Evaluating the Quality of Educational Programs in Pennsylvania: A Report From Educational Testing Service to the State Board of Education*, June 30, 1965. Harrisburg: State Board of Education, 1965.

Evaluative Criteria, 1960 Edition. Washington: National Study of Secondary School Evaluation, 1960.

Firman, William D., *et al*. *Procedures in School Quality Evaluation*. Albany: The University of the State of New York, the State Education Department, 1961.

Garber, Lee O. *The Law Governing School Board Members and School Board Meetings*. Danville, Illinois: Interstate Printers and Publishers, 1963.

Gross, Neal and Robert E. Herriott. *Staff Leadership in Public Schools: A Sociological Inquiry*. New York: John Wiley and Sons, Inc., 1965.

Grossman, Alvin and Robert L. Howe. *Data Processing for Educators*. Chicago: Educational Methods, Inc., 1965.

Hisch, Werner Z. "Education in the Program Budget," *Program Budgeting*, David Novick, editor. Washington: U. S. Government Printing Office, 1965. Ch. 6, pp. 131-54.

________, Elbert W. Sagelhorst, and Morton J. Marcus. *Spillover of Public Education Costs and Benefits*. Los Angeles: Institute of Government and Public Affairs, University of California, 1964.

Kershaw, J. A. and R. N. McKean. *Systems Analysis and Education*. Santa Monica, California: The Rand Corporation.

Krathwohl, David R., Behjamin S. Bloom, and Bertram B. Masia. *Taxonomy of Educational Objectives, Handbook II: Affective Domain*. New York: David McKay Company, Inc., 1964.

Leonard, John P., and A. C. Eurich. *An Evaluation of Modern Education*. New York: The Appleton Company, 1942.

Mort, Paul R., W. S. Vincent, and C. A. Newell. *The Growing Edge*. New York: Metropolitan School Study Council, 1957.

National Education Association Committee on Tax Education and School Finance. *Does Better Education Cost More?* Washington: The Association, 1959.

Office of Professional Development and Welfare, National Education Association. Profiles of Excellence. Washington: The Association, 1966.

Reutter, E. Edmund. The Administrator and Subversive Activities. New York: Bureau of Publications, 1951.

Shane, Harold G. and E. T. McSwain. Evaluation and the Elementary Curriculum. New York: Henry Holt and Company, 1951.

Smith, Eugene R., Ralph W. Tyler, and the Evaluation Staff. Appraising and Recording Student Progress. New York: Harper and Brothers, 1942.

Thorndike, Robert L., and Elizabeth Hagen. Measurement and Evaluation in Psychology and Education (Second Ed.). New York: John Wiley and Sons, Inc., 1961.

Tope, Donald E. The Social Science View of School Administration. Englewood, New Jersey: Prentice-Hall, 1965.

Travers, Robert M. W. Educational Measurement. New York: The MacMillan Company, 1955.

Trow, William Clark. Scientific Method in Education. Boston: Houghton Mifflin Company, 1925.

Tyler, Ralph W. Basic Principles of Curriculum and Instruction. Chicago: The University of Chicago Press, 1950.

Webb, Eugene J., et al. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally, 1966.

B. Periodicals and Unpublished Materials

Benson, Charles S. "Measurement of Quality in Education," *Science Education*, 50:283-5, April, 1966.

Boykin, L. L. "What is Evaluation?" *Progressive Education*, 34:16-18, January, 1957.

Blanchard, B. E. "Meaning of Evaluation in Education," *Bulletin of The National Association of Secondary School Principals*, 40:39-45, May, 1956.

Brumbaugh, A. J. "Role of Evaluation in the Development and Improvement Of Regional Programs of High Education," *Journal of Experimental Education*, 25:43-8, September, 1956.

Coleman, James S., *et al*. U. S. Office of Education, "Equality of Educational Opportunity," 1966.

Cook, Desmond L. *An Introduction to PERT*. Columbus, Ohio: Bureau of Educational Research and Service. Ohio State University, 1964. (Mimeographed.).

Coon, Lewis H. *Applying PERT to a Variable Manipulating Project*. Columbus, Ohio: Bureau of Educational Research and Service, Ohio State University, 1964. (Mimeographed.).

Dilley, Frank B. "Program Budgeting in the University Setting," *Educational Record*, Fall 1966.

Dyer, Henry S. "The Pennsylvania Plan," *Science Education*, 50:242-8, April, 1966.

Ebel, Robert L. "Some Measurement Problems in a National Assessment of Educational Progress," *Journal of Educational Measurement*, 3:11-17, Spring 1966.

Exton, Elaine. "Federal Program Budgeting Is a Step Toward Centralized Educational Planning," *American School Board Journal*, November, 1966.

———. "Here's How HEW Department Applies Planning-Programming-Budgeting System," *American School Board Journal*, December, 1966.

———. "USOE Uses Computer-Based Models to Evaluate Education," *American School Board Journal*, January, 1967.

Firman, William D. "The Quality Measurement Project in New York State," *Science Education*, 50:259-79, April, 1966.

______. "The Relationship of Cost to Quality in Education." A paper presented to the Committee on Educational Finance, National Education Association, St. Louis, Missouri, April 8, 1963. (Mimeographed.).

______. "Which Schools are Better?" NEA Research Bulletin, 41:83-9, October, 1963.

Flanagan, John C. "Evaluating Educational Outcomes," *Science Education*, 50:248-51, April, 1966.

Frymier, Jack R. "Curriculum Assessment: Problems and Possibilities," *Educational Leadership*, 24:124-8, November, 1966.

Gerberich, J. Raymond. "Assessment: A Forward Look," *Educational Leadership*, 24:115-8, November, 1966.

"Good School Value to Student Doubled," *The Houston Post*, Janaury 2, 1967.

Hagen, Elizabeth P. and Robert L. Thorndike. "Evaluation," *Encyclopedia of Educational Research* (Third Ed.) New York: The MacMillan Company 1960, pp. 482-6.

Hand, Harold C. "National Assessment Viewed as the Camel's Note," *Phi Delta Kappan*, 67:8-13, September, 1965.

Harris, Chester W. "The Appraisal of a School: Problems for Study," *Journal of Educational Research*, 41:172-83, November, 1947.

Hastings, J. Thomas. "Curriculum Evaluation: The Why of the Outcomes," *Journal of Educational Measurement*, 3:27-32, Spring, 1966.

Hollister, Robinson G., Jr. "A Decision-Making Budget," *Educational Record*, Fall, 1966.

"How to Evaluate Federal Programs," *Nation's Schools*, 77:5, May, 1966.

Howe, Harold II. "Growth and Growing Pains," *Saturday Review*, 49:68-70, 87, December 17, 1966.

Howsam, Robert B. "Teacher Evaluation: Facts and Folklore," *National Elementary Principal*, 43:7-18, November, 1963.

Journal of Educational Measurement, 3:1-65, Spring, 1966, (entire issue).

Kropp, R. P. "Evaluation Promotes Understanding of the Total School Program," *School Review*, 63:446-7, November, 1955.

Krooker, E. W. "Standards Versus Evaluation," Educational Administration and Supervision, 41:385-9, November, 1955.

Maccia, George S. An Educational Theory Model: General Systems Theory. Columbus, Ohio: Bureau of Educational Research and Service, Ohio State University, 1962. (Mimeographed.).

McDonald, Arthur S. "Some Pitfalls in Evaluating Progress in Reading Instruction," Phi Delta Kappan, 65:336-8, April, 1964.

Merwin, Jack C. "The Progress of Exploration Toward A National Assessment of Educational Progress," Journal of Educational Measurement, 3:5-10, Spring, 1966.

Mood, Alexander M. "Measurement of Quality in Education," Science Education, 50:279-83, April, 1966.

Nichols, Robert C. "Schools and the Disadvantaged," Science, 154:1312-1314, December 9, 1966.

Raths, L. E. "Basis For Comprehensive Evaluation," Educational Research Bulletin, 17:57-84, March, 1938.

______. "Role of Evaluation in Research Design," Educational Leadership, 13:412-14, April, 1956.

______. "Evaluating the Program of a School," Educational Research Bulletin, 17:57-84, November, 1938.

Ryan, T. Antoinette. "Testing Instructional Approaches for Increased Learning," Phi Delta Kappan, 66:534-6, June, 1965.

Shaftel, F. R. "Evaluation For Today and For the Future," Educational Leadership, 14:292-8, February, 1957.

Shane, Harold B. and Edward T. McSwain. "Evaluation of the Educational Program," Review of Educational Research, 23:171-80, October 1953.

Smallenberg, C. J. "Sound Evaluation," Educational Leadership, 21:11-16, October, 1963.

Smith, Dora V. "Recent Procedures in the Evaluation of Programs in English," Journal of Educational Research, 38:262-75, March, 1944.

Stufflebeam, Daniel L. Applying PERT to a Test Development Project. Columbus, Ohio: Bureau of Educational Research and Service, Ohio State University, 1964. (Mimeographed.).

Thomas, J. Alan. "Efficiency in Education." Paper Presented to American Education Research Association Meeting, Chicago, February 15, 1963. (Mimeographed.).

Tyler, Ralph W. "Assessing the Progress of Education," Phi Delta Kappan, 67:13-18, September, 1965.

_______. "Assessing the Progress of Education," Science Education, 50: 239-42, April, 1966.

_______. "The Objectives and Plans for a National Assessment of Educational Progress," Journal of Educational Measurement, 3:1-4, Spring, 1966.

_______. "Defining and Measuring Objectives of Progressive Education," Educational Research Bulletin, 15:67-71, 1936.

_______. "Modern Aspects of Evaluation," California Journal of Secondary Education, 29:410-412, 1954.